Universidad Carlos III de Madrid
www.uc3m.es

TESIS DOCTORAL

Essays on Forecasting with Partial Least Squares Methods

Autor:

Julieta Fuentes

Director/es:

Esther Ruiz Ortega

Pilar Poncela Blanco

Julio Rodríguez Puerta

DEPARTAMENTO DE ESTADÍSTICA

Getafe, Enero 2015

Universidad Carlos III de Madrid
www.uc3m.es

TESIS DOCTORAL

Essays on Forecasting with Partial Least Squares

Autor: **Julieta Fuentes**

Director/es: **Esther Ruiz Ortega**

Pilar Poncela Blanco

Julio Rodríguez Puerta

Firma del Tribunal Calificador:

Firma

Presidente: (Nombre y apellidos)

Vocal: (Nombre y apellidos)

Secretario: (Nombre y apellidos)

Calificación:

Getafe, de Marzo de 2015

Essays on Forecasting with Partial Least Squares Methods

A dissertation presented by

Julieta Fuentes

To
The Department of Statistics
In partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of

Business and Quantitative methods

Advisors:

Esther Ruiz
Universidad Carlos III de Madrid

Pilar Poncela and Julio Rodríguez Puerta
Universidad Autónoma de Madrid

Universidad Carlos III de Madrid
Octubre 2014

To Ale, Cami and Sofi

Acknowledgements

The development of this thesis has been a period of tremendous personal and professional growth for me. During the years of courses and thesis preparation that form the Doctoral program, I had the opportunity to meet many people to whom I am indebted, it is impossible for me to list them all, but without them, I would never have finished.

I would like to thank my supervisors Esther Ruiz, Pilar Poncela and Julio Rodríguez Puerta. To Esther Ruiz, for giving me the opportunity to develop my thesis and having proposed to Pilar Poncela as co-director. To Pilar, for blindly accepting to co-direct this thesis and for supporting me until the last minute. To Julio, for adding his invaluable counsel to this project. Pilar and Julio thanks for your orientation, for your patience, for all the work, time, experience and knowledge sharing; without you it would not have been able to complete this thesis, nor have my first journal publication. I will be eternally grateful to you.

I also want to express my gratitude to my family for the support they provided me through my doctoral studies. To my daughters, for motivating me every day to be a better person and enduring my absences, especially to Ale who lived with me the entire process, suffering in the early years all the changes that this implied. To my husband, for letting me grow and for having absorbed some of my family responsibilities. To my mother for her unconditional support. To my father by giving me the possibility of completing this thesis and to my aunt Ana Hilda for always being willing to help me.

I am also thankful to my fellow students, with whom I shared several experiences during the courses. I especially thank Mariana, Adolfo, Ana Laura, Ana María, Zulma and Alejandro. I also want to thank Don Narciso Bastero for helping me to settle in Madrid and my old friends Noemy and Miguel for encouraging me to keep going. As well, I want to thank Mr. Dennis Miller for “kindly, kindly” reviewed my English in a selfless manner.

Finally, I thank Central Reserve Bank of El Salvador for their support through the granting of work permits, as well as for participation in international conferences on forecasting.

“Dentro de nosotros hay algo que no tiene nombre,
esa cosa es lo que somos”.

José Saramago, 1995

Contents

Chapter 1

Introduction

In many fields of research the forecasting task involves a data rich framework, where the number of predictors N is considerable large to apply standard statistical techniques. In consequence, the development of new approaches to deal with these large informational sets has received increasing attention. Early theoretical and empirical research focused on dimension reduction methods, e.g. factor based models, however recent research has been extended to regularization methods. This thesis addresses the problem of construction of macroeconomic models for prediction from large databases through the application of techniques of dimension reduction and regularization, applied separately or simultaneously.

Principal components regression (PC) has been established as the most widely factor method used when high dimensional datasets are available. However, this approach presents some limitations. Two of the major criticisms towards PC and other dimension reduction techniques are the following: (i) when forming the linear combinations, the estimated factors do not take into account the target variable and (ii) the predictors may not have enough informative content about the target. In order to address the first PC weakness and thus consider the relationship between the predictors and the target variable, several approaches such as ad hoc rules (Boivin and Ng, 2006), targeted predictors (Bai and Ng, 2008) and partial least squares (Groen and Kapetanios, 2008) have been proposed. Furthermore, with the purpose of overcoming the second drawback a variety of selection methods have been introduced in the literature (Boivin and Ng, 2006; Bair et al., 2006; Bai and Ng, 2008; De Mol et al., 2008, among others), although some of them preserve the first limitation.

In this thesis we look at partial least squares (PLS), a technique which constructs a scheme for extracting orthogonal unobserved components based on the covariance between the predictors and the forecasting variable. Nonetheless, PLS methods are based on all variables, that is, it gives weight to all the predictors in the dataset; and then, the weight given to predictors with high predictive power is weakened. Therefore, we introduce into the economic analysis the sparse partial least squares (SPLS) approach, proposed by Chun and Keles (2010) in the context of chemometrics. In order to obtain a sparse solution, that is, to construct the factors based on a reduced number of predictors, this method includes an L1 penalty in the PLS formulation. This type of regularization allows selecting the predictors that contain relevant information and discarding those that have redundant information or negligible effect on the forecasting target variable. We extend the static implementation to a dynamic one, with the aim of considering the macroeconomic time series properties.

We consider three different lines of research for the SPLS regularization method: (i) to forecast a target variable y_{t+h} given an available set of information up to time t, (ii) to forecast simultaneously a set of variables Y_{t+h} given an available set of information up to time t and (iii) to build a combined forecast $y_{t+h/t}$ from the available multiple forecast of the same variable, with information up to time t. Consequently, the nature of the dataset employed to deal with these objectives is also different. In the first two cases, we used a large set of macroeconomic series (Stock and Watson database, 2005) and in the third one we considered the survey of professional forecasters (SPF) in which our focus is on experts point forecasts.

This thesis is structured as follows. In the second chapter we provide a literature review of the most widely used methods on forecasting with large data sets and on forecast combination, this latter focused on combining individual forecasts not models. In the third chapter, we revisit PLS method and discuss its dynamic implementation for macroeconomic series; furthermore, we suggest the use of the SPLS method, a technique that obtains dimension reduction and variable selection simultaneously, to forecast macroeconomic variables. Using the well-known Stock and Watson database, we explore the SPLS forecasting performance through a comparative exercise in which the most frequent methods used in the literature are considered: PC, targeted predictors (TP), Least Absolute Shrinkage and Selection Operator (LASSO) and PLS. In the fourth chapter we extend the PLS and SPLS univariate approach to the multivariate case. We explore if it is possible to take advantage of relations between a set of response variables to extract specific jointly factors (PLS2 and SPLS2), with the aim of obtaining better forecasting results. We enlarge the standard VAR model with PC, PLS2 and SPLS2 factors (VARF) and compare its forecasting performance. In the fifth chapter, the objective of regularization methods is directed toward the construction of a method to combine selected single forecasts from economic surveys in order to improve the forecast accuracy. We propose the use of the SPLS as a combination scheme. We employ the Survey of Professional Forecasters (SPF) dataset to explore the performance of different methods for combining forecasts: average forecasts, trimmed mean, Ordinary Least Squares (OLS), PLS, LASSO and SPLS. Finally, chapter six presents some conclusions and areas for further research.

Main Contributions

- We provide a literature review on the main methods currently used for macroeconomic forecasting with large dimensional data sets and on the most used approaches for forecast combination.
- We revisit partial least squares (PLS) and introduce Sparse Partial Least Squares (SPLS) into the economic analysis. The SPLS method has been used in chemometrics in a static context but has not been previously used for macroeconomic forecasting. We also propose its dynamic extension as in the case of PLS. The proposed methodology helps to overcome two of the major criticisms of factor methods: (i) the estimated factors do not depend directly on the prediction purpose and (ii) the available predictors may not have enough informative content about the target variable.
- We find that the choice of a useful or informative subset of predictors to extract the latent variables to forecast a specific target variable is relevant for improving the performance of the factor forecasting methods. The empirical comparison conducted among the forecasting performance of PLS, SPLS and the most widely used methods: PC, TP and LASSO for some variables of the Stock and Watson's database such as Consumer Price Index (CPI), Industrial Production (IP), Total employment (EMT), among others, show a good prediction performance of the SPLS model. This latter method is able to improve the forecast efficiency of the alternative methods.
- We introduce multivariate PLS and SPLS for macroeconomic forecasting, extending the idea of dimension reduction to forecast a group of variables rather than a single one, which is the usual practice.
- We find that the construction of jointly specific factors from a subset of relevant predictors is able to improve the forecast performance in the multivariate case. The VARF models augmented with multivariate PLS2, SPLS2 factors outperform the standard VAR models, and the VARF models augmented with PC factors.

- We find, through a restricted version of the VARF model, that the dynamics of a series is better capture only by means of its own lags.
- We propose the use of a regularized partial least squares (SPLS) as a method to combine selected single forecasts from economic surveys. The proposed methodology helps to find a weighting scheme that combines the available forecasts, considering the outcome and discarding the redundant and uninformative information, in order to generate forecasts with a better performance.
- We find that combination schemes that perform a forecasters selection yield to predictive gains over the widely used summarizing measure, the simple average of the survey participants. The empirical comparison performed among different methods for combining forecasts (average forecasts, trimmed mean, OLS, PLS and LARS) for some variables of the SPF shows that performing a forecasters selection improves the forecast efficiency compared with the simple average benchmark and that the selection process implemented by the SPLS method has a good prediction performance.

References

Bai, J. and S. Ng (2008). “Forecasting Economic Time Series Using Targeted Predictors”. *Journal of Econometrics*, 146: 304-317.

Bair, E., Hastie, T., Paul, D. and Tibshirani, R. (2006). “Prediction by Supervised Principal Components”. *Journal of the American Statistical Association*, 101: 119-137.

Boivin, J. and S. Ng (2006). “Are More Data Always Better for Factor Analysis?”. *Journal of Econometrics*, 132:169-194.

Chun H. and S. Keles (2010). “Sparse Partial Squares Regression for Simultaneous Dimension Reduction and Variable Selection”. *Journal of the Royal Statistical Society.* Series B, Statistical Methodology, 72(1): 3-25.

De Mol, C., Giannone, D. and L. Riechlin (2008). “Forecasting Using a Large Number of Predictors. Is Bayesian Regression a Valid Alternative to Principal Components?”. *Journal of Econometrics,* 146 (2): 318-328.

Groen, J. and G. Kapetanios (2008). “Revisiting Useful Approaches to Data-Rich Macroeconomic Forecasting”, Federal Reserve Bank of New York Staff Report 327.

Stock, J. and M. Watson (2005). “Implications of Dynamic Factor Models for VAR Analysis”. NBER Working Paper No. 11467.

Chapter 2

Literature Review

2.1 Introduction

The growing availability and access to large data sets and the facilities to process information provided by information technology have generated the opportunity to produce forecasts that are more accurate. However, it has also involved a challenge to standard forecasting techniques ("curse of dimensionality problem") which has led to the development of a number of alternative approaches for dealing effectively with all potentially relevant information for forecasting a specific target. For macroeconomic and financial variables several methods for forecasting with many predictors have been suggested. The most studied ones have been factor models, whether static or dynamic (see Stock and Watson, 1999, 2002a, 2002b; Forni et al., 2000, 2005; Kapetanios and Marcellino, 2009; among others).

The usefulness of factor models for the construction of economic indicators, policy analysis and forecasting has motivated several extensions within and outside this framework. The better understanding of the main aspects that affect the estimation of the factors and its employment for prediction: (i) size of the data set (Watson, 2000; Stock and Watson, 2004; Boivin and Ng, 2006; among others), (ii) quality and characteristics of the information e.g., oversampling and non-stationary variables (Schumacher, 2007; Schneider and Spitzer, 2004; Boivin and Ng, 2006, Kapetanios and Marcellino, 2009; Bai and Ng, 2004; among others), (iii) factor estimation technique e.g., supervised or unsupervised models (Boiving and Ng, 2005; Bai and Ng, 2008a), (iv) presence of structural breaks (see, for instance, Stock and Watson, 2009 and 2012a; Banerjee et al., 2008; Breitung and Eickmeier, 2011; Chen et al., 2014); among others has led to a development of different techniques to compute the factors and the emergence and introduction, into the economic analysis, of methods suited to exploit large data sets.

The theoretical and empirical literature on the analysis of large data sets has increased significantly in recent years, becoming a very active field of research in economics and related areas. Alternative methods such as forecast pooling (Figlewski, 1983; Stock and Watson 2003, 2004; Aiolfi and Timmermann, 2006), factor augmented VAR (Giannone et al., 2004; Bernanke et al., 2005), Bayesian model averaging (Koop and Potter, 2004; Stock and Watson, 2006), Bayesian regression (De Mol et al., 2008), Bayesian VAR (Banbura et al., 2010; Carriero, et al., 2009, 2012; Koop, 2013), least absolute shrinkage and selection operator (Tibshirani, 1996; Bai and Ng, 2008a) and bootstrap aggregation (Inoue and Kilian, 2008) have also been proposed.

The aims of this chapter are twofold. First, to provide a review of the literature on forecasting with large data sets, describing the most commonly used approaches and the development of factor models. Second, to provide a review of the literature on forecast pooling or forecast combination, with a focus on combining estimated individual forecasts, not models. Other surveys with empirical and/or theoretical emphasis can be found in Stock and Watson (2006, 2011 and 2012b), Eickmeier and Ziegler (2008), Bai and Ng (2008b), Eklund and Kapetanios (2008), among others. For forecast combination see, for instance, Timmermann (2006) and Aiolfi et al. (2011).

2.2 Forecasting Methods for Large Datasets

We distinguish two groups of forecasting methods for large data sets: (i) dimension reduction methods and (ii) regularization methods. The estimation of these methods requires, in general, two steps. In the first step a reduction or regularization process is performed e.g., the predictors are summarized by means of a factor estimation or a pre-selection of variables is made as a function of some particular decision rule and, in the second, the previous estimated factor or selected variables are used to construct a linear forecasting equation. The specification of the prediction model will depend on the approach used. In the case of traditional factor estimation, the prediction equation is given by:

$$y_{t+h} = \mu + \phi(L) y_t + \beta'(L) \hat{F}_t + \eta_{t+h} \quad (2.1)$$

where y_{t+h} is the target variable to be forecasted as a function of its own lags $\phi(L)\ y_t$ and of the factors and their lags estimated in the previous step $\beta'(L)\hat{F}_t$. $\beta(L)$ and $\phi(L)$ are lag polynomials of appropriate dimensions. The h-step ahead prediction error is denoted by η_{t+h}.

2.2.1 Dimension Reduction Methods

Dimension reduction methods constitute a way to overcome the curse of dimensionality problem when high dimensional data are present. Dimensionality reduction is the transformation of high dimensional data into a new lower dimensional dataset. Factor models have been considered among the methods that effectively reduce data dimensionality.

The main idea behind factor models is that variables can be represented as the sum of two mutually orthogonal unobservable components: (i) the common component, driven by few common factors to all variables in the model and (ii) the idiosyncratic one being composed of variable specific shocks. Let y_{t+h} be the target variable to be forecasted and X_t be a vector of candidate predictors of y_{t+h} at time t, of dimension (Nx1). Assuming that each predictor in X_t admits a factor structure as follow:

$$X_t = \Lambda F_t + \varepsilon_t \tag{2.2}$$

where F_t is an rx1 vector of unobserved common factors, ε_t is an Nx1 vector of disturbances and Λ is an Nxk coefficient matrix of factor loadings. Under the static representation, the common component (χ_t) is defined as ΛF_t. The factors and the idiosyncratic disturbances are assumed to be uncorrelated at all leads and lags, that is, $E(f_t \varepsilon_{is})=0$ for all i, s. A factor model with orthogonal idiosyncratic elements is called a strict factor model while an approximate factor model relaxes this assumption and allows a limited amount of correlation among the idiosyncratic terms (Chamberlain and Rothschild, 1983).

Several methods allow the estimation of approximate factor models when N is large and the factors are stationary. According to the meta-analysis performed by Eickmeier and Ziegler (2008), the two methods that dominate in the literature are the static and dynamic principal components proposed by Stock and Watson (2002b) and Forni et al. (2005), respectively. The authors pointed out a third, less widely applied, suggested by Kapetanios and Marcellino (2003).

The maximum likelihood estimation of the dynamic factor model for small models (N is considered small and finite) has been known for a long time in the literature (see, for instance, Geweke, 1977; Geweke and Singleton, 1981; Engle and Watson, 1981; for early contributions to this literature). However, recent results given by Jungbacker et al. (2011) and Doz et al., (2012) allow the estimation of large N dynamic factor models by maximum likelihood using the state space framework and Kalman filter techniques for large N models.

Moreover, Doz et al., (2012) show that the common factor estimates are consistent even though there is weak cross correlation in the error term not taken into account in the estimation procedure.

The dimension reduction approaches differ in a variety of assumptions, however they can be grouped in two broad types depending on whether the linear combination matrix is constructed taking into account the target variable or not. If the estimated factors are not directly related with the outcome, then the method will be classified as unsupervised, otherwise it will be considered as supervised.

2.2.1.1 Unsupervised Techniques

Static Principal Components

Principal Component Analysis (PC) constitutes the most popular dimension reduction approach. Stock and Watson (2002a) model the covariability of a large number of predictor series (N) in terms of a small number of unobserved latent factors, and they build forecasts using a linear regression between these estimated latent factors and the variable to forecast. Assume that X_t admits a factor model representation as in equation (2.2). The estimation of the factors is performed using the first k principal components of $\{X_t\}_{t=1}^{T}$, which are obtained by solving the following minimization problem in $\tilde{\lambda}_i$ and $\tilde{F}_t$

$$\min V = \min \sum_{i=1}^{N} \sum_{t=1}^{T} (X_{it} - \hat{X}_{it})^2 = \min \sum_{i=1}^{N} \sum_{t=1}^{T} (X_{it} - \tilde{\lambda}_i \tilde{F}_t)^2. \quad (2.3)$$

The solution of equation (2.3) provides the approximation with minimum mean square error for the X matrix. The problem is usually rewritten as the maximization of $N\, tr(\tilde{\Lambda}' X' X \tilde{\Lambda})$ subject to the identification restriction $\tilde{\Lambda}'\tilde{\Lambda}=I_r$ where tr (·) denotes the matrix trace. The objective is then to find the maximizer vectors ($\hat{\Lambda}$) of the diagonal sum of $\tilde{\Lambda}'\Sigma_{xx}\tilde{\Lambda}$, which is solved setting $\hat{\Lambda}$ equal to the eigenvectors of X′X corresponding to its k largest eigenvalues. The estimator of the factor is then constructed as $\hat{F}_t = \hat{\Lambda}' X_t$, the vector consisting of the first k principal components of X_t.

Under the identifying assumptions, several authors have shown that principal components consistently estimate the space spanned by the true common factors as N, T $\rightarrow \infty$ (Stock and Watson, 2002a; Forni et al., 2005; Bai and Ng, 2002).

The model proposed by Stock and Watson (2002a) is relatively simple to apply. Boivin and Ng (2005) showed that the factor model based on static principal components is quite robust to misspecification since very few auxiliary parameters have to be specified. However, it does not exploit the dynamics of the common factors.

Weighted Principal Components

Some authors such as Forni et al. (2005) and Innoue and Kilian (2008) have suggested other weighting schemes. Forni et al. (2005), assume that ε_t is i.i.d. N $(0, \Sigma_{\varepsilon\varepsilon})$ and propose weighting the predictors according to the inverse idiosyncratic variances, their signal to noise ratio. The authors estimate the idiosyncratic covariance matrix in the frequency domain as the difference between the sample covariance matrix of X_t and an estimator of the spectrum of the common component. Then, from the estimated covariance matrix of the common component, they solve a generalized eigenvalue problem and estimate $\hat{F}_t = \hat{\Lambda}_g ' X_t$, where $\hat{\Lambda}_g$, is the matrix whose columns are the first r largest generalized eigenvectors associated to the couple of matrices of the common component and idiosyncratic component. In practice the out off-diagonal elements of $\hat{\Sigma}_{\varepsilon\varepsilon}$ are set to zero (D'Agostino and Giannone, 2006). Thus, the estimated factors can be seen as static weighted principal components where the weights are derived from the estimated idiosyncratic covariance matrix.

Innoue and Kilian (2008) propose the employment of bootstrap aggregation or bagging for forecasting with large datasets. In particular, they introduce a method named bagging factor predictors. The method involves defining a pre-test estimator in the factor models, generating a large number of bootstrap samples between the target variable and the r largest common factors extracted by PC and compute the conditional bootstrap pre-test predictor for each sample. Then, the bagged predictor is estimated averaging the forecasts from the models selected by the pre-test on each bootstrap sample

2.2.1.2 Supervised Techniques

Factor models constitute an effective way to summarize the information contained in large datasets. For predictive purposes, it allows incorporating the information of all the predictors, but the estimated linear combinations are not necessarily the best. One of the major criticisms of the standard PC method is that it does not take into account the target variable in the factor construction step. Then, the estimated factors, which explain most variation in the predictors, may not explain the most variation in the variable to be forecasted. There have been some techniques suggested that attempt to overcome this drawback, among them, partial least squares (PLS) and reduced rank regression (RRR).

Partial Least Squares (PLS)

PLS is a dimension reduction technique, originally proposed by Wold (1966). This technique has been applied extensively in chemometrics and other related scientific areas but it has been less applied in economic literature (De Jong, 1993; Groen and Kapetanios, 2008). PLS attempts to address the unsupervised estimation of the factors in traditional PC, instead of searching for linear combinations as a function of the predictors covariance, it searches for linear combinations that maximize the covariance between the predictors and the target variable to construct the factors.

There are several algorithms available to estimate PLS factors; but, in general, the components are determined sequentially. The first PLS component is obtained as follows: $\hat{F}_1^{pls} = \alpha \sum_{i=1}^{N} Cov(x_{it}, y_{t+h}) x_{it}$, where α is a normalization constant. To compute the rest of components, the residuals of the regressions of each variable and the target over the preceding component are used. For example, to construct the second PLS component regress y_{t+h} and x_{it} on $\hat{F}_1^{pls}$ and a constant and let the residuals be $y_{1,t+h}^{pls}$ and $x_{1,it}^{pls}$, i=1, …, N then, obtain the component as follows: $\hat{F}_2^{pls} = \alpha \sum_{i=1}^{N} Cov(x_{1,it}^{pls}, y_{1,t+h}^{pls}) x_{1,it}^{pls}$. Then, the PLS estimated factors are orthogonal.

Reduced Rank Regression (RRR)

RRR is a multivariate linear regression method subject to a rank restriction on the coefficient matrix. It can be stated as a generalized eigenvalue problem, where the solution can be achieved by estimating the largest r eigenvalues of a symmetrical matrix formed by the product of the predictor covariance matrix, the responses covariance matrix and the covariance matrix of the predictors and the response. In this case, the factors are estimated taking into account the response variable or variables. The rank condition is equivalent to the number of components of PC selected to be included in the forecasting equation. If the rank condition is not considered, the estimation becomes a standard OLS problem.

2.2.2 Regularization Methods

According to Ng (2013), any method that prevents overfitting data is a form of regularization. For forecasting, empirical results do not support the basic principle that more data always improve statistical efficiency. Boiving and Ng (2006) pointed out that it is not simply N that influences the factor estimation and its forecast efficiency but the quality of the information. In fact, there is a growing body of evidence that indicates that the number and the properties of the variables included in the datasets are relevant for the factor estimation process (Bai and Ng, 2002; Boivin and Ng, 2006; Watson, 2000; Stock and Watson, 2009; Eickmeier and Ng, 2009; among others). Since the factor space being estimated is a function of the chosen panel of the predictor variables, the information content on the data is crucial to improve forecasting accuracy.

Two types of penalties have been introduced to prevent overfitting. The squared L2 norm implies weighting decay in the variables of the dataset according to their properties. In this case, all the available predictors are used. Ridge Regression, Bayesian regression and other ad hoc schemes (Boivin and Ng, 2006) are some of the methods proposed that shrink the matrix of predictors. The second type of penalty is the L1 norm, which tends to produce sparse models because it involves a variable selection process. Several variable selection approaches have been suggested, among them: Ad hoc schemes (Boivin and Ng, 2006; Bai and Ng, 2008a), supervised principal components (Bair et al., 2006), least absolute shrinkage selection (Hastie et al., 2008), sparse partial least squares (Chun and Keles, 2010) and Bayesian regression (De Mol, et al., 2008).

The resulting shrinkage and/or selection of an informative set of predictors can be used directly in the forecasting equation or can be used to construct factors, which in a next step will be introduced in the prediction model.

2.2.2.1 Down-weighting predictors

Ad hoc schemes

Different criteria have been implemented in the literature: size of idiosyncratic errors or amount in the error cross-correlation between variables and commonality ratio of each variable, among others. For example, Boivin and Ng (2006) employ weighting rules to account for two properties of the residuals: heteroskedasticity and cross correlation. The series with highly correlated errors or heteroskedasticity were down-weighted, but all series were used to estimate the factors. Caggiano et al. (2011) applied a modified version of the previous rules and others for the six largest euro area countries.

Ridge regression

The ridge regression is a commonly used method to solve the problem of multicollinearity. The ridge estimates are the solution of a penalized least squares criterion, with the penalty being proportional to the squared norm of the regression coefficient or vector β

$$\hat{\beta}_{\lambda} = \arg\min_{\beta}[\sum_{t=1}^{N}(y_t - \beta' x_t)^2 + \lambda\beta'\beta] \tag{2.4}$$

The solution to this problem is given by $\hat{\beta}_{\lambda} = (X^T X + \lambda I)^{-1} X^T y$. The tuning parameter λ controls the degree of regularization. If this is set equal to 0, then the ridge regression coefficients are the OLS estimators; while as $\lambda \to \infty$ the ridge coefficients tend to zero. The type of penalty introduced has the effect of shrinking the estimated coefficients towards zero. In the case of strongly correlated predictors, it shrinks the coefficients toward each other.

The scale of the variables is relevant for the solution of (2.4), and then, the variables need to be standardized before performing ridge regression.

2.2.2.2 Variable Selection

Ad hoc schemes

As in the case of down-weighting predictors several criteria have been employed to select variables from a large dataset: correlation coefficients between pair of errors and individual mean squared forecast error (MSFE), among others. Boivin and Ng (2006) introduced some rules based on the properties of the residuals to discard some variables from the panel and also performed a variable grouping forecasting exercise dividing the whole panel in three categories: real, nominal and volatile/leading indicators.

Bai and Ng (2008a) used a statistical test to determine whether a predictor will be in or out from the panel from which the factors will be estimated. In particular, the authors are interested in dropping the uninformative predictors. Therefore, to select the variables they employed a hard threshold procedure based on a statistical test to screen the variables from the dataset considered. In this case, the targeted predictors are selected as follows:

(a) A regression is performed between the variable to be forecasted (y_{t+h}) and each predictor variable; one constant and four lags of y_t are also included in the application.

(b) A threshold significance level α is set.

(c) A smaller set of predictors k_α^* whose t ratio defined as $|\hat{\beta}_i / se(\hat{\beta}_i)|$ exceeds the predefined threshold, are selected as targeted predictors.

(d) The factors are extracted by principal components from the reduced dataset (k_α^*).

(e) The forecast equation based on the previous extracted factors is estimated.

Supervised Principal Components (SPC)

SPC performs PCA using a subset of predictors that are chosen as a function of their association with the target variable. Bair et al. (2006) suggest the following procedure: (i) perform a univariate regression between each predictor and the target. SPC selects the variables with strongest estimated correlation with the response, (ii) form a subset of predictors with those that exceeds a predefined statistical threshold, (iii) perform PC using only the predictors selected in the previous step and (iv) estimate the forecasting equation with the first few PC factors computed in (iii).

Least Absolute Shrinkage and Selection Operator (LASSO)

LASSO is a penalized regression that performs shrinkage and variable selection. It was proposed by Tibshirani (1996) as a strategy to reduce the variance (sacrificing a little of bias) obtained by the OLS estimates. LASSO coefficients minimize a penalized residual sum of squares defined as follows:

$$\begin{aligned} &\hat{\beta}^{lasso} = \arg\min_\beta \sum_{i=1}^{n} (y_i - \sum_{j=1}^{p} x_{ij}\beta_j)^2 \\ &subject\ to \sum_{j=1}^{p} |\beta_j| \le \tau. \end{aligned} \qquad (2.5)$$

The tuning parameter τ controls the amount of shrinkage, if τ is sufficiently large then the constraint has no effect and the LASSO algorithm will yield the OLS estimates. However, if τ is sufficiently small, the constraint tends to produce some zero coefficients enforcing sparsity in the solution. The constraint is an L1 penalty and then some of the coefficients shrunk exactly to zero. Because there is no closed form expression for this problem several efficient algorithms have been developed for computing a solution such as least angle regression (LARS).

LARS can be considered similar to the forward stepwise regression algorithm, but proceeds as follows: first, it finds the predictor most correlated with the target, and then the coefficient of that predictor is increased (equiangularly), until that predictor is no longer the most correlated one with the residual. At this moment, a new variable is included in the active set. The selected variables can be included directly in the forecasting equation or can be used to estimate PC factors as a previous step. Ng (2013) highlighted that L1 penalties may improve mean-squared prediction accuracy if the bias-variance trade-off is favorable.

LASSO is closely related to ridge regression with a different type of penalty and to Bayesian regression under certain prior. The regularization process performed by Bayesian regression depends on the selection of the prior distribution, which is related to the L1 and L2 types of penalty, previously mentioned.

In the macroeconomic forecasting literature the most usual priors are the Gaussian and the double exponential. Under certain assumptions, the estimates under the Gaussian prior have been found to be equivalent to those produced by ridge regression, an L2 type of penalized regression (De Mol et al., 2008; Luciani, 2014). Thus, this prior corresponds to a case when all the predictors in the panel have a positive but decreasing weight. On the contrary, the double exponential prior favors either large or zero coefficients, being equivalent to an L1 penalized regression such as LASSO, under certain conditions. Consequently, its implementation results in variable selection. De Mol et al. (2008) analyzed the forecasting performance of these two prior distributions, considering that the regressors are linear combinations of the variables. The authors point out that these methods have barely been used for forecasting applications in large datasets. However, recent papers have extended their implementation to this data rich environment.

2.3 Forecast Combination

The idea of combining multiple individual forecasts to produce a pooled forecast was introduced by Bates and Granger (1969). The main argument for combining forecasts is taken from the classical diversification gains of the financial portfolio theory. Other reasons given in the literature for its application are the following: forecast combinations may serve as a hedge against the impact of structural breaks in individual forecasts, they can increase the forecast robustness by reducing the misspecification biases of models and the measurement errors in the data (associated to individual forecasts) and they can offset the individual forecasts bias generated by the use of different loss functions (Timmermann, 2006).

A variety of approaches have been proposed in the forecast combination literature. The research has been centered on the development of methods of estimating the combination weights (ω_{it}) in order to improve forecasting accuracy. Equal weighted average, least squares (optimal) estimators of the weights, relative performance weights, trimmed mean, Bayesian Model Averaging (BMA) and factor-based combinations are the most used schemes. Some of these methods can handle large datasets, however its empirical implementation has been limited (Stock and Watson, 2006; Conflitti et al., 2012). Recent extensions deal with issues such as weighting schemes that take into account possible structural breaks and combining density forecasts instead of point forecasts, among others (see, for instance, Wallis, 2011; Tian and Anderson, 2014).

We describe the most widely used weighting schemes when the number of predictors N can be large and when the input data are point forecasts, which comes from expert forecasters surveyed. In the first case, as was stated previously, there are at least two ways to address the problem: reduce the dimension or introduce some form of regularization. In the case of dimension reduction, in general, the combination is made from estimated factors (PC or PLS) or from forecasts estimated with blocks of predictors. BMA is also employed. The regularization is performed with methods such as LASSO. These methods may also be applied to survey forecast combination (Poncela et al., 2011; Conflitti et al., 2012; Genre et al., 2013). PC, PLS and LASSO were explained previously.

The goal is to forecast a target variable (y_{t+h}) from a combination of the multiple forecasts of the same variable with information up to time t: $y_{t+h} = \omega_0 + \sum_{i=1}^{N} \omega_{it} y_{i,t+h|t}$, where ω_{it} is the weight assigned to the ith individual forecast in period t.

Equal Weighted Average

The equally weighted average $\omega_{it} = 1/N$ is a widely used scheme for weighting the forecasts, which is equivalent to the simple arithmetic forecast. When N is large, ω_{it} may be very small and then, the informative predictors or experts may receive a low weight. However, empirical studies have shown that a simple equally weighted pooling forecast tends to outperform more sophisticated methods, which is known as the "forecast combination puzzle" (Stock and Watson, 2004; Timmermann, 2006; Genre et al., 2013).

Trimmed Mean

The trimmed mean is a simple combination forecast that involves trimming α% of the ordered observations from both ends. Thus, the trimmed mean is the mean after discarding some extreme values. The median represents the most trimmed statistics.

Least Squares (optimal) Weights

Granger and Ramanathan (1984) suggested estimating combination weights by ordinary least squares:

$$y_{t+h} = \omega_0 + \sum_{i=1}^{N} \omega_{it} y_{i,t+h|t} + \varepsilon_{t+h} \tag{2.6}$$

where the intercept term ω_0 is introduced to adjust any forecasts bias.

When N is large relative to T, combining using regression weights may be infeasible or inappropriate (sampling error); therefore, some type of dimension reduction has to be used in a previous step. However, Conflitti et al. (2012) suggested a strategy to compute optimal weights, when combining a large number of forecasts, adding additional constraints into the optimization problem.

Relative Performance Weights

Some alternative weighting schemes have been proposed in order to take into account the past forecasting performance of individual forecasts (subjective forecasts or models) in relation to the performance of the average model (Aiolfi and Timmermann, 2006; Stock and Watson, 2004, among others). For example, Aiolfi and Timmermann (2006) suggested constructing clusters of models based on their historical forecasting performance, computing a pooled average within each cluster and taking the average over models that pertain to the particular cluster or quartile of models that performed better than average forecast.

Stock and Watson (2004) introduced a weighting scheme where the weights depend inversely on the historical performance of each individual forecast. In particular, they used the discounted MSFE to build the combination: $\omega_{it} = m_{it}^{-1} / \sum_{j=1}^{n} m_{jt}^{-1}$, where

$m_{it} = \sum_{s=T_0}^{t-h} \delta^{t-h-s} (y_{s+h}^{h} - \hat{y}_{i,s+h|s}^{h})^2$ and δ is the discount factor. T_0 and t represent the beginning and end of the forecast period.

Bayesian Model Averaging (BMA)

The main idea behind this approach is to combine forecasts from a given set of models, computing the weights as formal posterior probabilities over the models. In a similar way than in Bayesian regression, the choice of priors can generate model selection. According to Timmermann (2006), these methods are increasingly used in empirical studies. Koop and Porter (2004) implemented BMA with factor-based models for forecasting GDP and the change of inflation.

2.4 Some Recent Extensions on Forecasting with Large Datasets

Maximum Likelihood Estimation

Peña and Poncela (2004) and Kapetanios and Marcellino (2003) proposed a state space model to estimating the factors from large datasets. The first case is applied in small panels possibly non-stationary, while the second case can handle large datasets and could be extended to deal with non-stationary factors. Recently Doz et al. (2012) have demonstrated that quasi-maximum likelihood estimation is feasible when N is large. The Kalman smoother and the EM algorithm allow computing the estimator. According to the authors, the estimator is a valid parametric alternative to PC. This method has some advantages such as the possibility of treating missing observations and extends the type of empirical applications that can be addressed.

Alternative Techniques for Dimension Reduction

Recent research has also been focused in alternative nonlinear dimension reduction techniques that are able to handle large datasets. Theoretically, models including nonlinearities may generate forecasting improvements. Among the suggested methods are independent component analysis (ICA) and sparse principal component analysis (SPCA) by Kim and Swanson (2014), random forest technique (Biau and D'Elia, 2009) and some kernel based methods: kernel ridge regression (Exterkate et al., 2013) and kernel principal component analysis (Giovannelli, 2012), among others.

Model Instability

Some research has been done in order to investigate the stability of the factor model forecasts. The conflicting results about the stability of the factor loadings and the influence of breaks on the number and estimation of factors have also led to the development of formal testing for structural breaks in this type of models (Stock and Watson, 2009 and 2012a; Banerjee et al., 2008; Breitung and Eickmeier, 2011; Chen et al., 2014).

2.5 Conclusions

The interest in harnessing the vast amount of information available has led to a growing amount of theoretical extensions and applications for forecasting, index modeling and policy analysis. In recent years, several methods for macroeconomic forecasting with large datasets have been proposed, some of them are new for the statistical literature and others are techniques developed in other areas of knowledge but new to econometrics. In this survey, we provide a brief description of the most widely used approaches for forecasting in a data rich environment, which has been organized with the aim of provide an overview of how it has evolved.

References

Aiolfi, M., Capistrán, C. and A. Timmermann (2011). Forecast Combinations. In *Oxford Handbook of Economic Forecasting,* Clements M. and Hendry, D. (eds). Oxford University Press; 355-388.

Aiolfi, M. and A. Timmermann (2006).“Persistence of Forecasting Performance and Combination Strategies”. *Journal of Econometrics*, 135 (1-2): 31-53.

Bai, J. and S. Ng (2002). “Determine the Number of Factors in Approximate Factor Models”. *Econometrica,* 70: 191-221.

Bai, J. and S. Ng (2004). “A Panic Attack on Unit Roots and Cointegration”. *Econometrica*, 72 (4): 1127-1177.

Bai, J. and S. Ng (2008a). “Forecasting Economic Time Series Using Targeted Predictors”. *Journal of Econometrics*, 146: 304-317.

Bai, J. and S. Ng (2008b). “Large Dimensional Factor Analysis”. *Foundations and Trends in Econometrics*, 3(2): 89–163.

Bair, E., Hasie, T., Paul, D. and Tibshirani, R. (2006). “Prediction by Supervised Principal Components”. *Journal of the American Statistical Association*, 101: 119-137.

Banbura M., Giannone D and L. Reichlin (2010). “Large Bayesian Vector Auto Regressions”. *Journal of Applied Econometrics,* 25: 71–92.

Banerjee, A., M. Marcellino, and I. Masten (2008). “Forecasting Macroeconomic Variables Using Diffusion Indexes in Short Samples with Structural Change”. *CEPR, Discussion Paper 6706*.

Bates, J. and C. Granger (1969). “The Combination of Forecasts”. *Operational Research Quarterly*, 20 (4): 451-468.

Bernanke, B., Boivin, J. and P. Eliasz (2005). “Measuring the Effects of Monetary Policy: a Factor-Augmented Vector Autoregressive (FAVAR) Approach”. *The Quarterly Journal of Economics,* 120(1): 387–422.

Biau, O. and A. D’Elia (2009). “Euro Area GDP Forecasting Using Large Survey Datasets. A Random Forest Approach”. Directorate General for Financial and Economic Affairs. European Commission.

Boivin, J. and S. Ng (2005). “Understanding and Comparing Factor-Based Forecast”. NBER Working Paper No. 11285.

Boivin, J. and S. Ng (2006). “Are More Data Always Better for Factor Analysis?”. *Journal of Econometrics*, 132:169-194.

Breitung, J. and S. Eickmeier (2011). “Testing for Structural Breaks in Dynamic Factor Models”. *Journal of Econometrics,* 163(1): 71–84.

Caggiano, G., Kapetanios, G. and V. Labhard (2011). “Are more Data Always Better for Factor Analysis? Results for the Euro Area, the Six Largest Euro Area Countries and the UK. *Journal of Forecasting*, 30(8): 736-752.

Carriero, A., Kapetanios, G. and M. Marcellino (2009). “Forecasting Exchange Rates with a Large Bayesian VAR”. *International Journal of Forecasting,* 25: 400–417.

Carriero, A., Kapetanios, G. and M. Marcellino (2012). “Forecasting Government Bond Yields with Large Bayesian VARs”. *Journal of Banking and Finance*, 36(7): 2026-2047.

Chamberlain, G. and M. Rothschild (1983). "Arbitrage, Factor Structure and Mean-Variance Analysis in Large Asset Markets. *Econometrica*, 51: 1281–2304.

Chen, L., Dolado, J. and J. Gonzalo (2014). "Detecting Big Structural Breaks in Large Factor Models". *Journal of Econometrics*, 180(1): 30-48.

Chun H. and S. Keles (2010). "Sparse Partial Squares Regression for Simultaneous Dimension Reduction and Variable Selection". *Journal of the Royal Statistical Society.* Series B, Statistical Methodology, 72(1): 3-25.

Conflitti, C., De Mol, C. and D. Giannone (2012). "Optimal Combination of Survey Forecasts". ECARES Working Paper 2012-023.

D'Agostino and Giannone (2006). "Comparing Alternative Predictors Based on a Large Panel Factor Models". Centre for Economic Policy Research, CEPR Discussion Papers No. 6564.

De Jong, S. (1993). "Simpls: An Alternative Approach to Partial Least Squares Regression". *Chemometrics Intelligent Laboratory Systems,* 18: 251–263.

De Mol, C., Giannone, D. and L. Riechlin (2008). "Forecasting Using a Large Number of Predictors. Is Bayesian Regression a Valid Alternative to Principal Components?". *Journal of Econometrics,* 146 (2): 318-328.

Doz, C., Giannone, D. and L. Riechlin (2012). "A Quasi Maximum Likelihood Approach for Large Approximate Dynamic Factor". *Review of Economics and Statistics*, 94 (4): 1014-1024.

Eickmeier, S. and C. Ziegler (2008). "How Successful are Dynamic Factor Models at Forecasting Output and Inflation? A Meta-Analytic Approach". *Journal of Forecasting*, 27: 237–265.

Eickmeier, S. and T. Ng (2009). "National Forecasting with Large International Datasets. An Application to New Zealand" Discussion Paper Series 11. Economic Studies from Deutsche Bundesbank, Research Centre.

Eklund J. and G. Kapetanios (2008). "A Review of Forecasting Techniques for Large Datasets" *National Institute Economic Review.*

Engle R. and M. Watson (1981). "A One Factor Multivariate Time Series Model of Metropolitan Wage Rates". *Journal of the American Statistical Association*, 76: 774-781.

Exterkate, P., Groenen, P., Heij, C. and D. Van Dijk (2013). "Nonlinear Forecasting with Many Predictors Using Kernel Ridge Regression". CREATES, Research Paper 2013-16. School of Economics and Management, University of Aarhus.

Figlewski, S. (1983). “Optimal Price Forecasting Using Survey Data”. *Review of Economic and Statistics* 65:813-836.

Forni, M., Hallin, M., Lippi, M. and L. Reichlin (2000). “The Generalized Dynamic Factor Model: Identification and Estimation”. *The Review of Economics and Statistics,* 82 (4): 540-554.

Forni, M., Hallin, M., Lippi, M., and L. Reichlin (2005). “The Generalized Dynamic Factor Model: One-Sided Estimation and Forecasting”. *Journal of the American Statistical Association,* 100 (471): 830–40.

Genre, V., Kenny, G., Meyler, A. and A. Timmermann (2013). “Combining Expert Forecast: Can Anything Beat the Simple Average?”. *International Journal of Forecasting*, 29:108-121.

Geweke, J. (1977). “The Dynamic Factor Analysis of Economic Time Series”, in *Latent Variables in Socio Economic Models,* edited by D.J. Aigner and A.S. Goldberger, Amsterdam: North Holland, 1-30.

Geweke J. and K. Singleton (1981). “Latent Variable Models for Time Series: A Frequency Domain Approach with an Application to the Permanent Income Hypothesis”. *Journal of Econometrics*, Vol. 17 (3): 287-304.

Giannone, D., Reichlin, L. and L. Sala (2004). “Monetary policy in real time”, NBER Macroeconomics Annual, 19:161-200.

Giovannelli, A. (2012). “Nonlinear Forecasting Using Large Datasets: Evidences on US and Euro Area Economies” CEIS Working Paper No. 255. University of Rome II.

Granger C. and R. Ramanathan (1984). “Improved Methods of Combining Forecast”. Journal of Forecasting, 3:197-204.

Groen, J. and G. Kapetanios (2008). “Revisiting Useful Approaches to Data-Rich Macroeconomic Forecasting”, Federal Reserve Bank of New York Staff Report 327.

Hastie T., Tibshirani, R. and J. Friedman (2008). *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer Series in Statistics, 2nd. edition.

Inoue, A. and L. Kilian (2008). “How Useful Is Bagging in Forecasting Economic Time Series? A Case Study of U.S. CPI Inflation”. *Journal of the American Statistical Association*, Vol. 103: 511 – 522.

Jungbacker, B., S.J. Koopman and M. van der Wel (2011). “Maximum Likelihood Estimation for Dynamic Factor Models with Missing Data”. *Journal of Economic Dynamics and Control*, 35: 1358 - 1368.

Kapetanios G. and M. Marcellino (2003). “A Comparison of Estimation Methods for Dynamic Factor Models of Large Dimensions”. Working Paper 489. Queen Mary University of London.

Kapetanios G. and M. Marcellino (2009). “A Parametric Estimation Method for Dynamic Factor Models of Large Dimensions”. *Journal of Time Series Analysis*, 30 (2): 208–238.

Kim, H. and N. Swanson (2014). “Mining Big Data Using Parsimonious Factor and Shrinkage”. ECB Workshop on using Big Data for Forecasting and Statistics, April.

Koop, G. (2013). “Forecasting with Medium and Large Bayesian VARS”. *Journal of Applied Econometrics*. 28: 177–203.

Koop, G. and S. Potter (2004). “Forecasting in dynamic factor models using Bayesian Model Averaging”. Econometrics Journal, 7:550-565.

Luciani, M. (2014). “Forecasting with Approximate Dynamic Factor Models: The Role of Non-Pervasive Shocks’. *International Journal of Forecasting*, 30: 20–29.

Ng S. (2013). “Variable Selection in Predictive Regressions”. *Handbook of Economic Forecasting,* Volume 2B. Graham Elliott, Clive Granger, Allan Timmermann (eds.), North Holland.

Poncela, P., Rodríguez, J., Sánchez-Mangas R. and E. Senra (2011). "Forecast Combination through Dimension Reduction Techniques". *International Journal of Forecasting,* 27: 224-237.

Peña, D. and Poncela, P. (2004). “Forecasting with Non-stationary Dynamic Factor Analysis. *Journal of Econometrics*, 119: 291-321.

Schneider, M. and M. Spitzer (2004). “Forecasting Austrian GDP using the Generalized Dynamic Factor Model”. Technical Report 89, Osterreichische Nationalbank.

Schumacher, C. (2007). “Forecasting German GDP using alternative factor models based on large datasets”. *Journal of Forecasting*, 26(4):271-302.

Stock, J. and M. Watson (1999). “Forecasting Inflation”. *Journal of Monetary Economics*, 44(2): 293-335.

Stock, J. and M. Watson (2002a). “Forecasting Using Principal Components from a Large Number of Predictors”. *Journal of the American Statistical Association*, 97(460): 1167-79.

Stock, J. and M. Watson (2002b). “Macroeconomic Forecasting Using Diffusion Indexes”. *Journal of Business & Economic Statistics*, 20: 147-163.

Stock, J. and M. Watson (2003). “Forecasting Output and Inflation: The Role of Asset Prices”. *Journal of Economic Literature,* 41:788-829.

Stock, J. and M. Watson (2004).“Combination Forecast of Output Growth in a Seven-Country Dataset”. *Journal of Forecasting*, 23:405-430.

Stock, J. and M. Watson (2006). “Forecasting with Many Predictors’. *Handbook of Economic Forecasting*, Graham Elliott, Clive Granger, Allan Timmermann (eds.), North Holland, 2006.

Stock, J. and M. Watson (2009). “Forecasting in Dynamic Factor Models Subject to Structural Instability”. In *The Methodology and Practice of Econometrics: Festschrift in Honor of D.F. Hendry*, Shephard and Castle (eds). Oxford: Oxford University Press.

Stock, J. and M. Watson (2011). “Dynamic Factor Models”. In *Oxford Handbook of Forecasting*, Michael P. Clements and David F. Hendry (eds). Oxford: Oxford University Press.

Stock and Watson (2012a), “Disentangling the Channels of the 2007-2009 Recession”. *Brookings Papers on Economic Activity*. Spring.

Stock, J. and M. Watson (2012b). “Generalized Shrinkage Methods for Forecasting Using Many Predictors”. *Journal of Business & Economic Statistics*, 30(4): 481-493.

Tian, J. and H. Anderson (2014). “Forecast Combinations under Structural Break Uncertainty”. *International Journal of Forecasting,* 30: 161–175.

Tibshirani, R. (1996). “Regression Shrinkage and Selection via the Lasso”. *Journal of the Royal Statistical Society*. Series B (Methodological), 58(1): 267-288.

Timmermann, A. (2006). “Forecast Combinations”. In *Handbook of Economic Forecasting*, Elliot, G., Granger, C. and A. Timmermann (eds.). North-Holland Vol I; 136-196.

Wallis, K. (2011). “Combining Forecasts – Forty Years Later”. *Applied Financial Economics*, 21: 33–41.

Watson, M. (2000). “Macroeconomic Forecasting Using Many Predictors”, In *Advances in Economics and Econometrics, Theory and Applications*, edited by M. Dewatripont, L. Hansen and S. Turnovsky, Eight World Congress of the Econometric Society, Vol. III: 87-115.

Wold, H. (1966). “Estimation of Principal Components and Related Models by Iterative Least Squares”. In *Multivariate Analysis,* edited by P.R. Krishnaiaah, New York: Academic Press, 391-420.

Chapter 3

Sparse Partial Least Squares in Time Series for Macroeconomic Forecasting

3.1 Introduction

Several approaches have been developed to deal with problems that are ill-posed due to the high dimensionality and multicollinearity of the information sets. In this chapter, with the aim of considering the forecast goal while reducing the dimension of the dataset, we revisit PLS and discuss its implementation, both static and dynamic, for time series in accordance with the properties of the data considered. Groen and Kapetanios (2008) is an early attempt in this line. The three-pass regression filter of Kelly and Pruitt (2012) can also be seen as a special case of PLS with one component.

Furthermore, considering the relevance of the quality of information contained in the dataset for forecasting (see, for instance, Bai and Ng, 2002; Boivin and Ng, 2006; Watson, 2000; Stock and Watson, 2009; Eickmeier and Ng, 2011; among others), instead of using all available predictors, we focus on the choice of a useful or informative subset of them to extract the latent variables and forecast a specific target variable. Note that the "factors", or to be more precise the unobserved common components, do not need to load in a great number of variables. On the contrary, the factor loadings can be zero for many predictors if they do not have enough informative content about the target variable. Bai and Ng (2008) is a first approximation to overcome this problem in the context of PC.

We introduce Sparse Partial Least Squares (SPLS) into the economic analysis, a technique that, besides taking into account the response variable for the component estimation, allows a variable selection process to be performed to construct a factor-forming subset. The SPLS method has been used in chemometrics in a static context (Chun and Keles, 2010). We also propose its dynamic extension as in the case of PLS.

Another feature of our proposal is that for stable relationships, it allows the best predictors to be selected within a large dataset, so it can be used as an exploratory tool. For unstable relationships, the re-estimation of the model allows the selection of the best predictors within the dataset based on their predictive content by monitoring the variables that go in/out of the model.

We use the Stock and Watson database in order to perform an empirical comparison of the forecasting performance of the PLS and SPLS methods to those widely used nowadays as principal components and targeted predictors. We focus our attention on forecasting inflation motivated by the reported difficulty to improve its performance, which is due in part to the changes in the inflation process and therefore in the instability of its predictive relationships (Stock and Watson, 2007 and 2009). The main findings confirm that there is some room for refinement in the factor forecasting methodology.

The chapter is organized as follows. Section 3.2 presents the forecasting framework. Section 3.3 discusses PLS in a dynamic context. Section 3.4 reviews some regularization methods already used in economics and introduces the sparse version of PLS (SPLS). We discuss its implementation, both static and dynamic. Section 3.5 presents the empirical application and provides the forecasting comparison for several horizons. Section 3.6 concludes.

3.2 Forecasting Framework

Our goal is to forecast y_{t+h} given the available information of the target up to time t, as well as from many other predictors, that we denote as X_t, and their lags. Since X_t can incorporate a large number of predictors, we would like to extract the information that is valuable for forecasting y_{t+h} in a parsimonious way. If the common information in X_t coincides with the useful information for forecasting y_{t+h}, we can use factor techniques to extract it. We use the term "factor" in a broad sense, meaning the unobserved component or signal that might be common to several variables (although not necessarily to many of them).

The forecasting model is specified and estimated as a linear projection of an h-step ahead transformed variable y_{t+h} onto t dated predictors. The predictors are the estimated factors, their lags and lags of the variable to be forecasted. That is,

$$y_{t+h} = \mu + \phi(L) y_t + \beta'(L) \hat{F}_t + \eta_{t+h} \tag{3.1}$$

where y_{t+h} is the variable to be forecasted at period t+h as a function of its own lags $\phi(L)\ y_t$ and of the factors and their lags estimated in the previous step $\beta'(L)\hat{F}_t$. The h-step ahead prediction error is denoted by η_{t+h}. The factor methods differ both in the way in which the factors are extracted and in the way in which the projection of the common component is made.

3.3 Partial Least Squares (PLS)

For the relationship between the target variable and the set of predictors when reducing the dimension, we use PLS. As stated before, the PLS orthogonal components are obtained iteratively. The first PLS factor $\hat{f}_{jt}^{PLS}$ can be computed from the eigenvalue decomposition of the matrix $M = X^{'}YY^{'}X$. To find the second PLS factor, the eigenvalue decomposition is performed on the residuals of the simple regressions of the target variable Y as well as on each of the predictors in X over the first PLS component. The process is repeated until the last factor has been extracted.

This technique has been implemented in a static way, while its implementation taking into account the dynamic behavior of the target is scarce (see Groen and Kapetanios, 2008; Eickmeier and Ng, 2011). We review the basic static application (static approach) and revisit and discuss how PLS can be applied to time series (dynamic approaches). We examine several types of approximations that account for the dynamics of the time series from alternative perspectives. The approaches differ in the set of predictors, the definition of the target adopted when extracting the factors and the estimation procedure.

To define the forecasting model, consider the following two equations:

$$y_{t+h} = \beta'(L) Z_t + \phi(L) y_t + u_{t+h} \tag{3.2}$$

and

$$Z_t = WX_t. \tag{3.3}$$

Equation (3.2) is our forecasting equation to produce the h-step ahead forecasts of the target variable, y. Forecasts are built as the sum of two components: their own dynamics collected in the term $\phi(L)y_t$; and the influence of the unobserved common components $\beta'(L)Z_t$. To highlight the difference between principal component regression and PLS, we denote the PLS components by $Z_t = \hat{f}_t^{PLS}$. The h-step ahead prediction error is denoted by u_{t+h}. Equation (3.3) expresses that the unobserved common components are formed as linear combinations of the candidate predictors through the weighting matrix W, where W is Nxk.

The key issue is how to estimate the unobserved components Z_t taking into account that we are dealing with time series. In what follows we discuss several static and dynamic possibilities.

Static Approach

a. The factors are extracted by applying PLS between the target variable (Y_{t+h}) and the original set of predictors (X). The lags of the target variable are included in the forecasting equation (3.2), whereas they are not taken into account when forming the unobserved common components Z_t. The M matrix is given by M= $X'Y_hY_h'X$, where $Y_h=(y_{h+1},\ldots,, y_{T+h})$ is the vector containing the target h periods ahead.

Dynamic approaches (DPLS)

b. The factors are based on applying PLS between an expanded set of predictors (X_e), enlarged with lags of the target variable, and the target variable (Y_h). The forecasting equation (3.3) does not include lags of the target, which are added as additional predictors in the linear combinations formed in equation (3.3).
c. The factors are based on applying PLS between the original set of predictors (X) and the residuals from an AR (p) process fitted to the target variable (Y_h). This can be done in a two step estimation procedure or in an iterative estimation algorithm. The lags of the target variable are included in the forecasting equation (3.2).

To illustrate the main ideas and see the potential advantages and disadvantages of each of the possibilities of applying PLS to time series data, we consider the simple case when the number of unobserved "factors" is k=1, the number of predictors is N=2 and the number of lags of the variable to be forecasted is just 1; so the AR filter in equation (3.2) is just $\phi(L)=\phi$, and we only need to include y_t. For the static approach (a), the h period ahead forecast is generated by the two step estimation of the following equations:

$$y_{t+h} = \beta_1 Z_t + \phi y_t + u_{t+h} \quad (3.4)$$

$$Z_t = w_1 x_{1t} + w_2 x_{2t} \quad (3.5)$$

where $Z_t = \hat{f}_t^{PLS}$ and w_i, i=1,2, are the weights assigned to each one of the predictor variables in the PLS component.

In a first step, the direction vector w is found by solving the following optimization problem:

$$w = \arg\max_w \; w' X_t' Y_{t+h} Y_{t+h}' X_t w \quad subject\ to\ w'w = 1 \quad (3.6)$$

with w=($w_1, \ldots, w_r)'$,which leads to the following objective function for the case r=2

$$\max_{(w_1 w_2)} w_1^2 [\sum_{t=1}^{T} x_{1t} y_{t+h}]^2 + 2 w_1 w_2 [\sum_{t=1}^{T} x_{1t} y_{t+h}][\sum_{t=1}^{T} x_{2t} y_{t+h}] + w_2^2 [\sum_{t=1}^{T} x_{2t} y_{t+h}]^2 + \lambda (w_1^2 + w_2^2 - 1)$$

where $[\sum_{t=1}^{T} x_{it} y_{t+h}]$ is T times the covariance between each predictor in X and Y_{t+h}. Solving the previous problem, we obtain that in the first PLS component the direction vector w, which is a function of the covariances between each of the predictors (X_t) and the target variable (Y_{t+h}):

$$w_i = \frac{\sum_{t=1}^{T} x_{it} y_{t+h}}{\sqrt{[\sum_{t=1}^{T} x_{1t} y_{t+h}]^2 + [\sum_{t=1}^{T} x_{2t} y_{t+h}]^2}} \quad for\ i = 1, 2. \quad (3.7)$$

In a second step, once the factor $\hat{f}_t^{PLS} = Z_t$ has been estimated, it enters equation (3.4) to serve as a reduced set of explanatory variables. The dynamic relationships of the target variable are captured directly through the inclusion of its own lags as explanatory variables in the forecasting equation.

In the dynamic approach (b), the model set up is as follows:

$$y_{t+h} = \beta_1 Z_t + u_{t+h} \tag{3.8}$$

$$Z_t = w_1 x_{1t} + w_2 x_{2t} + w_3 y_t . \tag{3.9}$$

In this case, instead of incorporating the lags of the target variable (Y_t) as regressors in the forecasting equation, they are included as additional predictors in X. In our simple illustration, the expanded data set contains three predictor variables, where $x_{3t}=y_t$. The direction vectors are estimated by solving the optimization problem (3.6), for r=3:

$$w_i = \frac{\sum_{t=1}^{T} x_{it} y_{t+h}}{\sqrt{[\sum_{t=1}^{T} x_{1t} y_{t+h}]^2 + [\sum_{t=1}^{T} x_{2t} y_{t+h}]^2 + [\sum_{t=1}^{T} y_t y_{t+h}]^2}} \quad for \ \ i = 1,2,3. \tag{3.10}$$

Notice that if PLS assigns a weight to all the variables included in the data set, the AR(p) process associated to the target variable (AR(1) in this simple setup) will attenuate its participation as $N\rightarrow\infty$. Then, if the AR(p) process is relevant for explaining the target variable, as is the case for macroeconomic variables, this approach could have a poor performance relative to the “static approach”, where the AR(p) process is included directly in the forecasting equation.

Approach (c) proposed an alternative way to integrate the dynamic relationship in the factor estimation that consists in isolating the effect of the AR(p) process before the PLS estimation. The forecasting framework can be expressed as in (3.4) and (3.5), but the optimization problem (3.6) is modified as follows:

$$w = \arg\max_w \ w' X_t' Y Y' X_t w \quad subject \ to \ w' w = 1 \tag{3.11}$$

where Y= [Y_{t+h}-ϕY_t]. In our example, the problem can be stated as:

$$\max_{(w_1 w_2)} [w_1 [\sum_{t=1}^{T} x_{1t}(y_{t+h} - \phi y_t)] + w_2 [\sum_{t=1}^{T} x_{2t}(y_{t+h} - \phi y_t)]]^2 + \lambda(w_1^2 + w_2^2 - 1).$$

The estimated direction vectors for this alternative method preserve the same structure as the previous ones but depend on the AR(1) coefficient:

$$w_i = \frac{\sum_{t=1}^{T} x_{it}(y_{t+h} - \phi y_t)}{\sqrt{[\sum_{t=1}^{T} x_{1t}(y_{t+h} - \phi y_t)]^2 + [\sum_{t=1}^{T} x_{2t}(y_{t+h} - \phi y_t)]^2}} \quad for \;\; i = 1,2 \tag{3.12}$$

where ϕ is the autoregressive coefficient of the AR(1) model that captures the target variable's own dynamics, before the PLS estimation.

3.4 Sparse Methods

There is a growing body of literature which suggests that the selection of relevant variables from a large feasible set is needed to improve forecast efficiency inside economics (i.e. Bai and Ng, 2008 propose TP and Dobrev and Schaumburg, 2012, apply regularization to reduced rank regressions) as well as outside economics (see, for instance, Lee et al., 2011 and Lê Cao et al., 2008 for the application of SPLS). Bai and Ng (2008) proposed forecasting economic series using a reduced set of informative variables named targeted predictors (TP). The last authors combine a variable selection process with PC estimation. Two types of threshold rules (hard and soft thresholds) are introduced in order to take into account the relation between the whole dataset and the variable of interest. For the hard threshold rule the authors used a statistical test to determine whether a predictor will be in or out from the panel from which the factors will be estimated. The detailed procedure was described previously as an ad hoc scheme for variable selection.

The authors used LASSO estimator as the soft thresholding rule. As was mentioned before LASSO is a shrinkage method in regression analysis that can be used to select covariates depending on the degree of shrinkage (see, Hastie et al., 2008). LARS is a model selection algorithm that approximates the first order conditions of the optimization problem solved in LASSO. Bai and Ng (2008) used this procedure to rank the predictors and then performed PC over the selected ones (soft thresholding). Instead, we will use LARS as a pure selection procedure and include in a regression framework the first k ranked predictors as opposed to combining them.

In order to overcome simultaneously the two drawbacks pointed out for extracting the "factors" (not taking into account the forecasting goal and too much uncertainty because weight is given to all predictors) Chun and Keles (2010) propose the SPLS static formulation in the context of biology.

The SPLS approach imposes an additional constraint (λ) on the PLS method, which operates on the direction vectors and leads to sparse linear combinations of the original predictors given in terms of a surrogate vector (c). They define a two objective optimization problem where the weights are defined by the θ parameter, which controls the effect of the concavity of the objective function and the closeness of the original vector (w) and the surrogate direction vector (c)

$$Min_{w,c} - \theta w' M w + (1-\theta)(c-w)' M (c-w) + \lambda \mid c \mid_1$$

(3.13) $$subjet \;\; to \; w' w = 1.$$

The additional term that appears in the optimization problem is given in terms of λ, the sparsity parameter, which is a penalty that encourages sparsity on the direction vector. When θ=1, the first term is the original eigenvalue problem of PLS if $M=X^{'}YY^{'}X$, and of PC if $M=X^{'}X$. When $M=X^{'}X$, the problem becomes the LASSO approach in PC analysis (SCoTLASS) when w=c and Sparse PC analysis (SPCA) when θ=1/2 (see Zou and Hastie, 2005).

Notice that SPLS can work through several components. This is one of the main differences with the TP approach of Bai and Ng (2008). The possibility of more than one SPLS component gives a chance to variables that are marginally significant to enter into the linear combinations that form the factors. Additionally, correlation among the predictors is not taken into account in TP.

We include SPLS into the economic analysis. In particular, we explore its usefulness in the macroeconomic forecasting area, so we consider $M=X^{'}YY^{'}X$. Since only its static version is available in the literature, we also consider its extension to the dynamic case. In fact, we apply this methodology for the same alternative approaches proposed for PLS, and include dynamics in the sparse version of PLS.

3.5 Empirical Application

To check how the different procedures perform in terms of forecasting accuracy, we use the Stock and Watson (2005) database and an updated version of it. The target variable is the US logarithm of the consumer price index (CPI), which is assumed to be integrated of order 2 (Stock and Watson, 2002; Bai and Ng, 2008) and is defined as:

$$y_{t+h}^{h} = \frac{1200}{\text{h}}(y_{t+h} - y_t) - 1200(y_t - y_{t-1}). \tag{3.14}$$

and let

$$z_t = 1200\,(y_t - y_{t-1}) - 1200\,(y_{t-1} - y_{t-2}). \tag{3.15}$$

The forecasting model is estimated at each period as a function of its own lags $\phi(L)\, y_t$ and the estimated factors (F_t) and their lags. The parameters and factors are estimated with information up to time t (X_t and y_{t+h-1}). The number of lags of the predictors is chosen by the Bayesian information criterion (BIC). We consider several forecast horizons h=1, 6, 12 and 24 to check the performance of the different approaches in the short and medium run. The final forecasts are obtained as follow:

$$y_{t+h}^{h} = \mu + \phi(L) z_t + \beta'(L)\hat{F}_t \tag{3.16}$$

It is important to state that instead of selecting some particular number of factors, derived from a particular criterion, we extract different number of factors from the data set and allow the final number of factors to be determined by the forecasting performance.

The original data set consists of 132 monthly United States (U.S.) macroeconomic time series that span the period from January 1960 through December 2003, for a total of T=528 observations .

The series are transformed to achieve stationarity by taking logs, first or second differences as necessary as in Bai and Ng (2008); Stock and Watson (2006). For data definitions and transformations, see Appendix A.

For comparison purposes, we employ seven forecast subsamples, as defined by Bai and Ng (2008), which can account for the temporal instability in the relation between the predictors and the variable to forecast. For factor estimation, the initial period of the dataset is always March 1960, whereas the final period is recursively expanded from February 1970 to February 1980 and February 1990 onwards until the end of each sample. The estimation and forecast samples are summarized in Table 3.1.

3.5.1 Forecast Results

The predictive ability of the PC, TP, LARS, PLS and SPLS methods over a univariate benchmark is compared in Tables 3.2 to 3.5 for the different forecast horizons considered. We use as benchmark an AR(4) for h=1. For the remaining forecast horizons, we also regress y_{t+h}^{h} over z_t and three lags. As the measure for forecast comparison, we use the relative mean-squared forecast errors (RMSE) over the benchmark:

$$RMSE\ (method) = \frac{MSE\ (method)}{MSE\ (AR(4))}. \tag{3.17}$$

An entry of less than one implies an improvement of the method upon the simple AR(4) forecast.

As regards PC regression, we try k=1 to 10 for the number of factors. Their lags, as well as the number of lags for the target variable, are selected by the BIC. We also borrow some of the forecasting results from Bai and Ng (2008). In particular, we consider the relative mean square forecast errors from the TP, in which the factors are estimated from a subset of the available data, using hard threshold rules and from the PC method where factors are estimated from the whole data set of predictors. We compare different forecasting methods against LARS that captures the idea of regression with pure (instead of combining) selection of variables. Following Bai and Ng (2008) we have decided to keep 5 and 10 variables.

The PLS approach is implemented in the different versions considered in section 3.3 in order to take into account the properties of the data. With the aim of evaluating the SPLS forecast performance, we estimate the latent SPLS components, considering values for the sparsity parameter λ in the set {0.2, 0.4, 0.6 and 0.8}. The number of components considered is k=1 and 2, although we have tried up to five components. Since the best forecasting results were obtained most of the time with just two components, we perform a more complete analysis for k=1 and 2.

Tables 3.2 to 3.5 show the forecasting results for forecasting horizons h=1, 6, 12 and 24. They suggest some interesting observations of the competing methods. First, the results highlight that it is possible to make refinements to the factor forecasting methodology. We find efficiency gains over the widely used PC and over PLS and LARS by estimating sparse factors predictors by TP and SPLS. Second, SPLS yields the most precise forecast in all the subsamples considered for the 24 months forecast horizon and in 71% of the subsample periods for the 1 month and 12 months forecast horizons. For the remaining subsamples, its accuracy is similar to the best alternative models. Third, in general, the improvements upon the benchmark are larger for longer forecast horizons. For instance, in Table 3.5 where h=24, SPLS gives the best forecasting results in all subsamples. The better forecasting performance at longer horizons has also been found in the factor model literature (see, for instance, Matheson, 2006; Caggiano et al, 2011).

Fourth, as regards the performance of PLS, it is important to note that when the dynamic relationship of the target variable is directly captured in the forecasting equation through the lagged values of the target variable (options a and c), the method provides better results, outperforming the benchmark and PC and TP too, except for the first sample. In the particular case of h=6, PLS outperforms the rest of the models in four samples. Nevertheless, when the lags of the forecasting variable are incorporated as additional predictors in the dataset (option b), the method performs even worse than the benchmark. The reason is that PLS gives weights to all the predictors, and then since the dimension of the cross section N is large, the weight given to z_t and its lags is weakened with respect to the options in which they are included directly in the forecasting equation. Note, however that this is not necessarily the case with its sparse version, where this option performs well in the short run (for h=1) in almost all samples, which can be seen from column 10 of Table 3.2. In the particular case of h=1, the lags of the variable have a large predictive power with respect to other explanatory variables. The selection process seems to weight appropriately the relevant information for the prediction purpose, disregarding variables that have a negligible effect on the response and enough weight is given to z_t and its lags in order to capture the dynamic behavior of y_{t+h}^{h}.

Fifth, the SPLS options a and c, produces the best results for h=12 in almost all samples, while for the remaining subsample its prediction accuracy is similar to the best PLS models.

Sixth, the SPLS, option a produces the best results for h=24 in all samples, which indicates that for these horizons some (but not all) the predictor variables, as distinct to the lags of the target, contain relevant information about the variable to forecast. The estimated

RMSE reported in Table 3.5 (column 9), SPLS (option a) outperforms PC, TP and LARS (5) and (10).

It seems that selecting variables to build the factors improves forecast accuracy. As a suggestion made by one of the referees, in order to formally compare the predictive accuracy of the competing models, we perform a pairwise comparison of all methods mentioned above against a pure selection procedure such as LARS (5) by computing the Diebold and Mariano test (Diebold and Mariano, 1995). The results for CPI are summarized in Table 3.6.

The implementation of the DM test suggests that SPLS forecast are significantly different from LARS (5). The LARS (5) model is dominated by a version of SPLS and/or PLS models in almost all the samples and horizons as can be seen from columns 4 to 9 of Table 3.6. We do not find significant differences between LARS performed selecting 5 or 10 variables but in one case out of the 28 analyzed.

3.5.2 The Variables Chosen

The average number of chosen variables is given in Table 3.7 (for h=1 and 6) and Table 3.8 (for h=12 and 24).

For h=1 and the best performing option in each model, the average number of chosen variables goes up to 88, when considering all the forecasting subsamples (see Table 3.7 columns three to five). The instability of the forecast period 1970 to 1980 seems to influence these results; when the forecast subsample does not include this period, the number of variables selected is significantly reduced and only goes up to 16, as shown in columns 3 through 5 of Table 3.7. The outcomes imply a high degree of sparsity if the unstable period of the 70's is not included.

The average number of variables chosen is higher for the remaining forecast horizons, probably due to growing uncertainties and the necessity to account for many sources of variability to explain the behavior of the target variable

As regards the variables chosen we can group them in components of CPI (especially when SPLS is performed through options a and c), unemployment, real activity variables such as components of IP and monetary variables as monetary bases (real M2) and the Fed funds rate. There are some variables related to demand, such as the consumption variable, consumption credit and the purchasing manager's index (PMI).

If we focus solely on h=1, when the lags of the target variable are included in the expanded set of predictors X_e, as happens in option b, the first lag of the CPI has a sizable contribution to forecasting CPI in the short term. For example, in the third forecasting sample that includes the period from 1990 to 2000, a stable decade of relatively mild inflation, the method selects only the first lag and the services price component as relevant variables. This phenomenon has been associated to improved monetary policy making, a result of smaller and more infrequent shocks hitting the economy and a structural break in the relationship between the inflation and the common factors, which constitute the most frequently explanations for The Great Moderation (Bernanke, 2004; Summers, 2005; Kim et al., 2004; among others). For the rest of samples, the predictors selected are the usual ones, price components, monetary aggregates and employment. The variables that dominate the list are CPI-U services, personal consumption expenditures deflactor on nondurables (PCE ND), exchange rate of Swiss franc, 5 years Treasury bond (TB) and the monetary base (MB).

The price components disappear as the forecast horizon h grows, while the remaining variables appear more frequently: unemployment, industrial production (IP) and its components, monetary aggregates and interest rates and, to a lesser extent, variables related to consumption, indicating more complex relationships between inflation and its possible sources of variation.

A comparison among the variables selected for LARS (5) and SPLS models pointed out that both select a similar category of variables for each horizon, but the frequency of choice of each type is distinct. For example, in the case of h=1 the two methods choose the price components as the most informative variables, but LARS (5) incorporates a larger number of series related to consumption than SPLS does. For h=6 there are some clear differences in the selection: LARS (5) model tends to select production and financial variables, interest rates and exchange rates, while SPLS concentrates in employment and monetary aggregates variables. Respect to the longer horizons, the emphasis on financial spreads by LARS (5) is the main discrepancy between the models. In both cases the employment, production variables, price components and real M2 are the most frequent selected variables.

3.5.3 Other Variables

To test the empirical validity of the sparse factor models, we apply the proposed procedures to the series: IP, total employment (EMT), personal income (PI) and retail sales (RS). The analyzed series in the chapter are those that appear on Bai and Ng's (2008) paper. These are in the spirit of the business cycle analysis (see http://www.nber.org/cycles/general_statement.html) and are in accordance with other business cycle analysis. See, for instance, the monthly indicator presented in Stock and Watson (1991). For the sake of brevity, and like Bai and Ng (2008), we report the results only for h=12 and assume that the log level of the four series are differenced stationary. The target variables are defined as follows:

$$y_{t+h}^{h} = \frac{1200}{\mathrm{h}}(y_{t+h} - y_t) \text{ and } z_t = 1200(y_t - y_{t-1}). \tag{3. 18}$$

The results are reported in Tables 3.9 to 3.12. We apply the methodology for k=1 and k= 2 components, given that in most cases the best forecasting results were obtained with a small number of components. The degree of sparsity is high in all samples and models (λ=0.8) with the exception of the first one. As mentioned for inflation, the instability of this forecast period seems to require more variables to account for the variance of the target.

We might draw the following conclusions from the tables: First, we find that PLS and SPLS methods provide the best forecasting results except for one subsample (70.3-80.12) for IP and one subsample (80.3-90.12) for RS in which TP gives the best results. Second, if we focus on SPLS we find forecast improvements with respect to TP in 93% of the cases. This ratio increases to 100% of the cases when we compared the performance of SPLS to that of LARS (5) and LARS (10). These results support the findings for CPI about the possibility of obtaining gains in terms of statistical accuracy by the employment of sparse factor models.

3.5.4 Updated Dataset

We perform an update of the Stock and Watson (2005) dataset. The updated base contains 112 monthly macroeconomic time series, and extends the time series of the original base through December 2010 for a total of T= 610 observations. For comparison purposes, we divide the updated dataset into the three subsamples shown in Table 3.13

The initial period for factor estimation is always March 1960, as in the previous application, whereas the final periods are recursively expanded from February 1970 and January 2000 until the end of the setup. As a reference, the first updated subsample coincides with the last one considered for the original database.

We implement the three different versions of the PLS and SPLS approaches proposed in section 3.3 and the standard Principal Components (PC (10)). Table 3.14 summarizes the forecasting results for h=1, 6, 12 and 24. The main findings regarding this update are the following: SPLS seems the best forecasting procedure for h=1, 6, 12 and 24, it produces the most accurate forecast in 83% of the samples, as columns 8 and 10 confirm it. Second, the improvements over the benchmark are larger for sample 2, which means that SPLS provides a better forecast for the 2000s decade. Third, SPLS gives better results for options a and c, where the dynamics of the target are taken into account in the forecasting equation rather than in the selection method.

Table 3.15 shows the average number of variables selected. As in the previous cases, when the forecasting sample includes the 70s, the number of variables is larger, while it is reduced by around 30 for the 21st century.

3.6 Conclusions

The empirical results are encouraging, suggesting that there is some room for refinement the factor forecasting methodology. The dynamic SPLS methodology introduced in this chapter shows a good prediction performance, improving the forecast efficiency of the alternative widely used factor methods in macroeconomic forecasting. Our findings confirm that the choice of a useful or informative subset of predictors, to extract the latent variables to forecast a specific target variable is relevant for improving the performance of the factor forecasting methods. More variables (more information) do not necessarily yield better forecasting results.

Among the different possibilities analyzed to apply PLS and SPLS to time series data, it seems that applying directly the PLS techniques between the target variable and the predictors yields the better forecasting results. Enlarging the data set of predictors, by including the lags of the target variable in it, does not seem to be a good alternative for PLS when applied to time series data, although this is not necessarily the case when the sparse version is applied. The PLS method gives weight to all the forecasting predictors, so the dependence between the target variable and its past can be obscured if there are too many predictors. Conversely, including the lags of the target variable explicitly on the forecasting equation seems to be the best way of capturing the dynamic behavior of the target.

The forecast performance of SPLS improves with the forecast horizon. This might reflect the fact that when the dynamics of the own lags die out, the predictive content of the cross section emerges. This is observed in most of the approaches analyzed in contrast to the pure AR(4). Taking into account that in the very short run (h=1), the forecasting results given by all the methodologies are much closer; the dynamic SPLS approaches seem to perform quite well. When the dynamic relationship is integrated through the inclusion of the lags of the target as additional predictors in the original dataset, the selection process seems to weight the relevant information for forecasting purposes appropriately. In particular, the presences of variables that have a negligible effect on the response do not lessen the participation of z_t and its lags. For the updated dataset, the isolation of the AR(p) process effects, before PLS estimation, shows also a good performance at all forecasting samples.

The variable selection performed by the SPLS model shows differences between the periods of high and low uncertainty in the economic environment and among the forecasting horizons and thus evidences the relevance of increasing the flexibility in the factor forecasting methodology. The proposed SPLS method has more flexibility than the traditional benchmarks; it allows choosing suitable predictors period by period to forecast a target and monitoring the variables that go in/out the model, so it can also be used as an exploratory tool.

Additionally, the variables chosen by the SPLS model in the CPI case have an economic foundation. The variables chosen to forecast inflation are mainly monetary variables, such as interest rates and monetary aggregates (real M2), price components and real activity variables, such as unemployment, housing starts and industrial sector activity indicators. There are some variables associated to the demand side, such as consumption and consumption credit, and sales (manufacturing and trade sales and retail sales). A greater interpretability of the results is an additional gain of the proposed methodology.

References

Bai, J. and S. Ng (2002). “Determine the Number of Factors in Approximate Factor Models”. *Econometrica,* 70: 191-221.

Bai, J. and S. Ng (2008). “Forecasting Economic Time Series Using Targeted Predictors”. *Journal of Econometrics*, 146: 304-317.

Bernanke, B. (2004). "The Great Moderation', Remarks by B. Bernanke Member of the Board of Governors of the U.S. Federal Reserve System, at the meetings of the Eastern Economic Association, Washington, D.C. and BIS Review 12/2004.

Boivin, J. and S. Ng (2006). "Are more Data Always Better for Factor Analysis?". *Journal of Econometrics*, 132: 169-194.

Caggiano, G., Kapetanios G., and V. Labhard (2011). "Are More Data Always Better for Factor Analysis? Results for the Euroarea, the Six largest Euroarea Countries and UK". *Journal of Forecasting*, 30 (8): 736-752.

Chun H. and S. Keles (2010). "Sparse Partial Squares Regression for Simultaneous Dimension Reduction and Variable Selection". *Journal of the Royal Statistical Society.* Series B, Statistical Methodology, 72(1): 3-25.

Diebold, F. X. and R. Mariano (1995). "Comparing Predictive Accuracy". *Journal of Business and Economic Statistics*, 13: 253-265.

Dobrev, D. and E. Schaumburg (2012). "Robust Forecasting by Regularization", Working Paper.

Eickmeier, S. and T. Ng (2011). "Forecasting National Activity using lots of International Predictors: An Application to New Zealand". *International Journal of Forecasting*, 27 (2): 496-511.

Groen, J. and G. Kapetanios (2008). "Revisiting Useful Approaches to Data-Rich Macroeconomic Forecasting", Federal Reserve Bank of New York Staff Report 327.

Hastie T., Tibshirani, R. and J. Friedman (2008). *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer Series in Statistics, 2nd. edition.

Kelly, B. and S. Pruitt (2012). "The Three-Pass Regression Filter: A New Approach to Forecasting Using Many Predictors", Chicago Booth Paper No. 11-19. Fama-Miller Center for Research in Finance.

Kim, C.J., Nelson, C. and J. Piger (2004). "The Less Volatile U.S. Economy: A Bayesian Investigation on Timing, Breadth, and Potential Explanations". *Journal of Business and Economic Statistics*, Vol. 22: 80-93.

Lee, D., Lee, W., Lee, Y. and Y. Pawitan (2011). "Sparse Partial Least-Squares Regression and its Applications to High-Troughput Data Analysis". *Chemometrics and Intelligent Laboratory Systems, 109: 1-8.*

Lê Cao, K.A., Rossouw, D., Robert-Granié, C. and P. Besse (2008). "A Sparse PLS for Variable Selection when Integrating Omics Data". *Statistical Applications in Genetics and Molecular Biology*, 7, Article-35.

Mathenson, T.D. (2006). “Factor Model Forecast for New Zealand”. *International Journal of Central Banking*, Vol. 2 (2).

Stock, J. and M. Watson (1991). “A Probability Model of the Coincident Economic Indicators”. In *Leading Economic Indicators: New Approaches and Forecasting Records,* edited by K. Lahiri and G. Moore, Cambridge University Press, p 63-90.

Stock, J. and M. Watson (2002). “Macroeconomic Forecasting Using Diffusion Indexes”. *Journal of Business & Economic Statistics*, 20: 147-163.

Stock, J. and M. Watson (2005). “Implications of Dynamic Factor Models for VAR Analysis”. NBER Working Paper No. 11467.

Stock, J. and M. Watson (2006). “Forecasting with Many Predictors.” In The *Handbook of Economic Forecasting*, Vol. 1, edited by G. Elliott, C. Granger and A. Timmermann, Elsevier Science, 516-542.

Stock, J. and M. Watson (2007). “Why U.S. Inflation Become Harder to Forecast”. *Journal of Money, Credit and Banking*, 39: 13-23.

Stock, J. and M. Watson (2009). “Forecasting In Dynamic Factor Models Subject To Structural Instability”. In *The Methodology and Practice of Econometrics, Festschrift in Honour of Professor David F. Hendry*, edited by J. Castle and N. Shephard, Oxford: Oxford University Press, p 1-57.

Summers, P. (2005). “What Causes the Great Moderation? Some Cross-Country Evidence”, *Economic Review*, Federal Reserve Bank of Kansas City. 3rd quarter.

Watson, M. (2000). “Macroeconomic Forecasting Using Many Predictors”, In *Advances in Economics and Econometrics, Theory and Applications*, edited by M. Dewatripont, L. Hansen and S. Turnovsky, Eight World Congress of the Econometric Society, Vol. III: 87-115.

Zou, H., T. Hastie (2005). “Regularization and Variable Selection Via the Elastic Net”. *Journal of Royal Statistical Society*, Series B, 67(2): 301-320.

Table 3.1
Estimation and forecast subsamples

SS	Estimation subsample	Forecast subsample
M1	1960:03 to 1970:03-h	1970:03 to 1980:12
M2	1960:03 to 1980:03-h	1980:03 to 1990:12
M3	1960:03 to 1990:03-h	1990:03 to 2000:12
M4	1960:03 to 1970:03-h	1970:03 to1990:12
M5	1960:03 to 1970:03-h	1970:03 to 2000:12
M6	1960:03 to 1980:03-h	1980:03 to 2000:12
M7	1960:03 to 1970:03-h	1970:03 to 2003:12

Table 3.2
RMSE, h=1

Period	Bai and Ng (2008)		LARS		PLS			SPLS		
	PC (10)	TP	5	10	a. (k=1)	b. (k=2)	c. (k=2)	a. (k=2)	b. (k=2)	c. (k=2)
70.3-80.12	1.015	**0.944**	1.102	1.190	1.054	1.365*	0.957*	1.022*	1.022*	0.958*
80.3-90.12	0.982	0.874	1.018	1.022	0.965	1.010*	0.917*	0.844	**0.819**	0.883
90.3-00.12	0.963	0.938	1.015	1.025	1.015	1.160	0.943	**0.772**	0.797	0.801*
70.3-90.12	0.998	**0.952**	1.067	1.112	1.022	1.201*	0.955*	0.974	1.012*	0.955*
70.3-00.12	0.990	0.954	1.059	1.098	1.019	1.222*	0.951	**0.939**	0.987	0.952*
80.3-00.12	0.972	0.895	1.019	1.025	0.975	1.071	0.921*	0.844	**0.809**	0.862
70.3-03.12	0.979	0.945	1.047	1.092	1.014	1.229*	0.943	**0.921**	0.979	0.943*

Source: Bai and Ng (2008) and authors' calculations. The table shows the ratio of MSE of PC, TP, PLS and SPLS over the benchmark model for h=1. PLS and SPLS results are shown only for the number of components k which yields the best forecasting results. An asterisk (*) means k=1.Bold figures indicate the best forecasting method for each subsample.

Table 3.3
RMSE, h=6

Period	Bai and Ng (2008)		LARS		PLS			SPLS		
	PC (10)	TP	5	10	a. (k=2)	b. (k=2)	c. (k=2)	a. (k=2)	b. (k=2)	c. (k=1)
70.3-80.12	0.712	0.713	0.786	0.719	**0.498**	1.056	0.548	0.508	0.890*	0.552
80.3-90.12	0.654	0.647	0.789	0.794	0.562*	1.005	0.574*	**0.530***	0.871	0.556
90.3-00.12	0.660	0.639	0.986	1.066	0.574*	1.187	0.614*	**0.573***	0.761	0.609
70.3-90.12	0.675	0.672	0.815	0.789	**0.567**	1.044	0.591	0.571	0.890*	0.587
70.3-00.12	0.671	0.677	0.825	0.810	**0.558**	1.048	0.582	0.561	0.895*	0.578
80.3-00.12	0.652	0.656	0.808	0.826	0.548*	1.018	0.564*	**0.525***	0.853	0.551
70.3-03.12	0.670	0.678	0.817	0.803	**0.561**	1.053	0.587*	0.565	0.899*	0.583

Source: Bai and Ng (2008) and authors´ calculations. The table shows the ratio of MSE of PC, TP, PLS and SPLS over the benchmark model for h=6. An asterisk (*) means k=1. Bold figures indicate the best forecasting method for each subsample.

Table 3.4
RMSE, h=12

Period	Bai and Ng (2008)		LARS		PLS			SPLS		
	PC (10)	TP	5	10	a. (k=2)	b. (k=2)	c. (k=2)	a. (k=2)	b. (k=2)	c. (k=2)
70.3-80.12	0.631	0.652	0.606	0.554	0.462	0.890	0.472	0.472	0.765	**0.458**
80.3-90.12	0.575	0.565	0.641	0.710	0.486	0.995	0.476	**0.440**	0.784	0.441
90.3-00.12	0.723	0.616	1.032	0.989	0.532	1.200	0.621*	**0.521**	0.820*	0.623
70.3-90.12	0.603	0.608	0.624	0.626	0.465	0.947	0.465	0.469	0.780	**0.465**
70.3-00.12	0.611	0.613	0.670	0.666	**0.466**	0.957	0.473	0.470	0.823	0.473
80.3-00.12	0.594	0.581	0.717	0.765	0.484	1.005	0.487	**0.461**	0.858	0.473
70.3-03.12	0.609	0.616	0.680	0.671	**0.469**	0.977	0.479	0.472	0.845	0.478

Source: Bai and Ng (2008) and authors´ calculations. The table shows the ratio of MSE of PC, TP, PLS and SPLS over the benchmark model for h=12. An asterisk (*) means k=1. Bold figures indicate the best forecasting method for each subsample.

Table 3.5
RMSE, h=24

Period	Bai and Ng (2008)		LARS		PLS			SPLS		
	PC (10)	TP	5	10	a. (k=2)	b. (k=2)	c. (k=2)	a. (k=2)	b. (k=2)	c. (k=2)
70.3-80.12	0.532	0.501	0.539	0.542	0.271	0.657	0.305	**0.266**	0.480	0.300*
80.3-90.12	0.506	0.545	0.535	0.545	0.372	0.769	0.375	**0.339**	0.628	0.360
90.3-00.12	0.546	0.626	0.975	0.767	0.464*	1.273	0.487*	**0.454***	1.253	0.475
70.3-90.12	0.522	0.531	0.537	0.547	0.321	0.710	0.342	**0.302**	0.545	0.339
70.3-00.12	0.523	0.538	0.572	0.564	0.332	0.752	0.353	**0.315**	0.600	0.349
80.3-00.12	0.512	0.558	0.599	0.576	0.384	0.837	0.391	**0.358**	0.716	0.376
70.3-03.12	0.523	0.538	0.574	0.565	0.334	0.759	0.356	**0.318**	0.606	0.352

Source: Bai and Ng (2008) and authors´ calculations. The table shows the ratio of MSE of PC, TP, PLS and SPLS over the benchmark model for h=24. An asterisk (*) means k=1. Bold figures indicate the best forecasting method for each subsample.

Table 3.6
Diebold and Mariano Test

Period		PLS			SPLS		
	10	a.	b.	c.	a.	b.	c.
h=1							
70.3-80.12	-1.159	0.675	-1.377	1.404	1.270	1.480	1.349
80.3-90.12	-0.121	0.788	0.081	1.253	2.142	**2.760**	1.642
90.3-00.12	-0.242	0.007	-1.191	0.795	**2.575**	2.935	2.893
70.3-90.12	-1.017	0.865	-1.190	1.590	1.459	1.415	1.558
70.3-00.12	-1.039	0.889	-1.662	1.801	**2.165**	1.180	1.748
80.3-00.12	-0.192	0.806	-0.696	1.559	2.971	**3.713**	2.398
70.3-03.12	-1.327	0.771	-2.024	1.862	**2.404**	1.189	1.824
h=6							
70.3-80.12	1.637	**1.845**	-2.191	1.260	1.853	-0.621	1.236
80.3-90.12	-0.213	2.403	-1.115	2.434	**2.899**	-0.499	2.323
90.3-00.12	-2.081	2.358	-1.084	1.992	**2.312**	1.429	1.984
70.3-90.12	0.964	**2.587**	-1.790	2.072	2.616	-0.525	2.118
70.3-00.12	0.584	**2.993**	-1.898	2.432	3.033	-0.556	2.475
80.3-00.12	-0.762	2.955	-1.257	2.899	**3.247**	-0.327	2.725
70.3-03.12	0.599	**3.071**	-2.114	2.464	3.109	-0.922	2.503
h=12							
70.3-80.12	1.536	1.654	-2.474	1.488	1.717	-3.312	**1.412**
80.3-90.12	-2.114	1.120	-2.871	1.415	**2.002**	-0.993	2.223
90.3-00.12	0.553	2.734	-0.716	2.105	**2.673**	1.415	2.104
70.3-90.12	-0.045	1.920	-3.678	2.117	1.856	-1.985	**2.113**
70.3-00.12	0.159	**2.621**	-3.035	2.755	2.557	-2.033	2.744
80.3-00.12	-1.283	1.941	-2.101	2.197	**2.663**	-1.117	2.786
70.3-03.12	0.343	**2.844**	-3.310	2.958	2.785	-3.079	2.957
h=24							
70.3-80.12	-0.293	1.972	-1.042	1.785	**1.944**	0.423	1.760
80.3-90.12	-0.376	1.353	-2.378	1.491	**1.627**	-1.011	1.849
90.3-00.12	1.884	2.534	-1.212	2.329	**2.587**	-0.976	2.159
70.3-90.12	-0.723	2.292	-2.160	2.228	**2.468**	-0.100	2.356
70.3-00.12	0.442	2.664	-2.279	2.610	**2.794**	-0.335	2.737
80.3-00.12	0.776	2.003	-2.270	2.142	**2.209**	-1.322	2.490
70.3-03.12	0.489	2.671	-2.361	2.603	**2.796**	-0.376	2.730

Note: A t-statistics greater than 1.65 means Ho is rejected at the 10% level of significance and a value greater than 1.96 denotes rejection at the 5% level of significance. Bold figures indicate the best forecasting method for each subsample.

Table 3.7
Average number of selected variables. h=1 and h=6

Period	h=1				h=6			
	TP	SPLS			TP	SPLS		
		a. (k=2)	b. (k=2)	c. (k=1)		a. (k=2)	b. (k=1)	c. (k=1)
70.1-80.1	32.170	3.446	1.000	85.323	43.562	111.585	6.023	79.823
80.1-90.1	62.421	16.385	2.177	8.962	70.868	34.415	1.000	12.131
90.1-00.1	73.884	6.515	2.000	1.000	72.132	80.331	1.000	66.508
70.1-90.1	47.299	7.428	1.000	88.104	57.195	111.760	5.336	79.384
70.1-00.1	56.152	7.119	3.438	87.514	62.158	111.459	4.578	78.581
80.1-00.1	68.154	6.684	2.092	7.552	71.494	30.944	1.000	12.468
70.1-03.9	57.757	7.017	3.314	86.672	63.003	110.983	1.000	78.457

Source: Bai and Ng (2008) and authors´ calculations. The table shows the average number of selected variables with the different methods applied to form the linear combinations that constitute the unobserved "factors".

Table 3.8
Average number of selected variables. h=12 and h=24

Period	h=12				h=24			
	TP	SPLS			TP	SPLS		
		a. (k=2)	b. (k=2)	c. (k=2)		a. (k=2)	b. (k=2)	c. (k=2)
70.1-80.1	64.223	36.523	24.223	63.369	73.132	62.939	19.323	75.646
80.1-90.1	87.264	15.677	11.231	13.508	92.769	83.354	27.746	39.692
90.1-00.1	89.000	78.915	1.000	102.585	96.430	113.815	68.008	31.208
70.1-90.1	75.714	110.316	10.308	98.928	82.917	73.016	23.420	34.320
70.1-00.1	80.127	111.581	8.438	100.184	87.418	76.592	24.000	33.368
80.1-00.1	88.124	15.684	8.020	11.936	94.606	83.684	26.528	35.704
70.1-03.9	80.634	111.570	32.746	100.148	87.418	77.205	24.003	32.956

Source: Bai and Ng (2008) and authors´ calculations. The table shows the average number of selected variables by each method for h=12 y h=24.

Table 3.9
RMSE, IP, h=12

Period	Bai and Ng (2008)		LARS		PLS			SPLS		
	PC (10)	TP	5	10	a. (k=2)	b. (k=2)	c. (k=1)	a. (k=2)	b. (k=2)	c. (k=2)
70.3-80.12	0.247	**0.197**	0.421	0.457	0.270	0.955	0.423	0.225	0.940	0.386
80.3-90.12	0.846	0.820	0.699	0.745	**0.544**	0.925	0.718	0.546	0.894	0.715
90.3-00.12	1.055	1.327	0.839	0.832	**0.549**	0.989	0.888	0.574	0.988	0.892
70.3-90.12	0.442	0.399	0.519	0.585	0.383	0.947	0.472	**0.368**	0.932	0.449
70.3-00.12	0.497	0.483	0.576	0.638	0.435	0.952	0.483	**0.421**	0.938	0.474
80.3-00.12	0.898	0.944	0.833	0.895	**0.690**	0.958	0.818	0.692	0.949	0.820
70.3-03.12	0.551	0.526	0.637	0.683	0.486	0.956	0.461	0.472*	0.943	**0.451**

Source: Bai and Ng (2008) and authors´ calculations. The table shows the ratio of MSE of PC, TP, PLS and SPLS over the benchmark model for h=12. PLS and SPLS results are shown only for the number of components k that yields the best forecasting results. An asterisk (*) means k=1.

Table 3.10
RMSE, EMT, h=12

Period	Bai and Ng (2008)		LARS		PLS			SPLS		
	PC (10)	TP	5	10	a. (k=2)	b. (k=2)	c. (k=1)	a. (k=2)	b. (k=2)	c. (k=2)
70.3-80.12	0.524	0.487	0.581	0.581	0.356	0.522	0.526	**0.286**	0.509	0.505
80.3-90.12	0.644	0.656	0.781	0.851	**0.483**	0.665	0.647	0.485	0.665	0.664
90.3-00.12	0.947	0.965	1.206	1.131	0.869	0.801	0.808	0.866	**0.795**	0.805
70.3-90.12	0.569	0.549	0.689	0.713	0.423	0.551	0.560	**0.389**	0.548	0.562
70.3-00.12	0.616	0.601	0.761	0.772	0.481	0.558	0.573	**0.451**	0.555	0.571
80.3-00.12	0.730	0.744	0.913	0.945	0.663	0.796	0.802	**0.659**	0.796	0.813
70.3-03.12	0.696	0.689	0.819	0.823	0.549	0.527	0.537*	**0.520**	0.523	0.533

Source: Bai and Ng (2008) and authors´ calculations. The table shows the ratio of MSE of PC, TP, PLS and SPLS over the benchmark model for h=12. PLS and SPLS results are shown only for the number of components k that yields the best forecasting results. An asterisk (*) means k=2. Bold figures indicate the best forecasting method for each subsample.

Table 3.11
RMSE, PI, h=12

Period	Bai and Ng (2008)		LARS		PLS			SPLS		
	PC (10)	TP	5	10	a. (k=1)	b. (k=1)	c. (k=1)	a. (k=1)	b. (k=1)	c. (k=1)
70.3-80.12	0.545	0.465	0.634	0.756	0.477	0.475	0.447	0.444	**0.432**	0.443
80.3-90.12	0.902	0.944	0.946	0.935	0.644	0.633	0.702	0.633	**0.626**	0.732
90.3-00.12	1.106	1.096	0.945	0.953	0.916	0.900	0.984	0.862	**0.858**	0.945
70.3-90.12	0.673	0.635	0.782	0.861	0.607	0.603	0.625	0.587	**0.583**	0.607
70.3-00.12	0.796	0.766	0.825	0.894	0.687	0.680	0.721	0.674	**0.667**	0.724
80.3-00.12	1.012	1.027	0.950	0.961	0.773	0.760	0.843	0.770	**0.760**	0.884
70.3-03.12	0.817	0.782	0.862	0.894	0.690	0.684	0.726	0.676	**0.669**	0.738

Source: Bai and Ng (2008) and authors´ calculations. The table shows the ratio of MSE of PC, TP, PLS and SPLS over the benchmark model for h=12. PLS and SPLS results are shown only for the number of components k that yields the best forecasting results. Bold figures indicate the best forecasting method for each subsample.

Table 3.12
RMSE, RS, h=12

Period	Bai and Ng (2008)		LARS		PLS			SPLS		
	PC (10)	TP	5	10	a. (k=2)	b. (k=1)	c. (k=2)	a. (k=1)	b. (k=1)	c. (k=2)
70.3-80.12	0.620	0.633	0.654	0.751	0.498	0.507	0.502	**0.417**	0.432	0.427
80.3-90.12	0.559	**0.514**	0.758	0.750	0.528	0.560	0.552	0.522	0.528	0.525
90.3-00.12	1.158	1.142	1.229	1.041	1.016	**0.985**	1.025	0.997	0.998	0.994
70.3-90.12	0.601	0.601	0.719	0.774	0.529	0.551	0.542	**0.478**	0.504	0.496
70.3-00.12	0.716	0.713	0.824	0.835	0.635	0.644	0.647	**0.602**	0.611	0.617
80.3-00.12	0.840	0.808	0.935	0.870	0.721	0.725	0.739	0.725	0.721	**0.706**
70.3-03.12	0.726	0.723	0.829	0.840	0.647	0.659*	0.660	0.625*	**0.622**	0.626

Source: Bai and Ng (2008) and authors´ calculations. The table shows the ratio of MSE of PC, TP, PLS and SPLS over the benchmark model for h=12. PLS and SPLS results are shown only for the number of components k that yields the best forecasting results. An asterisk (*) means k=2. Bold figures indicate the best forecasting method for each subsample.

Table 3.13
Estimation and forecast subsamples

SS	Estimation subsample	Forecast subsample
M1	1960:03 to 1970:03-h	1970:03 to 2003:12
M2	1960:03 to 2000:02-h	2000:02 to 2010:12
M3	1960:03 to 1970:03-h	1970:03 to 2010:12

Table 3.14
RMSE, CPI

Period	PC	LARS		PLS			SPLS		
h=1	10	5	10	a. (k=1)	b. (k=1)	c. (k=1)	a. (k=1)	b. (k=1)	c. (k=1)
70.3-03.12	0.979	1.053	1.108	1.014	1.233	0.946	0.989	0.998	**0.942**
00.2-10.12	0.809	1.395	1.763	0.730	1.038	0.799	**0.713**	0.901	0.740
70.3-10.12	**0.905**	1.230	1.440	0.936	1.136	0.917	0.942	0.986	0.908
h=6				a. (k=2)	b. (k=2)	c. (k=2)	a. (k=2)	b. (k=1)	c. (k=2)
70.3-03.12	0.670	0.829	0.784	0.584	1.059	0.605*	0.582	0.899	**0.587***
00.2-10.12	0.969	1.225	1.388	0.836*	1.018	0.702	**0.592**	0.668	0.674
70.3-10.12	0.791	1.035	1.094	0.718	1.035	0.795*	**0.606**	0.883	0.750
h=12				a. (k=2)	b. (k=2)	c. (k=2)	a. (k=2)	b. (k=2)	c. (k=2)
70.3-03.12	0.609	0.661	0.672	0.480	0.977	0.489	0.482	0.821	**0.491**
00.2-10.12	0.709	0.513	0.498	0.530*	1.335	0.496	**0.417***	0.665*	0.485
70.3-10.12	0.605	0.597	0.598	0.491*	1.103	0.564*	**0.471***	0.753	0.549
h=24				a. (k=2)	b. (k=2)	c. (k=2)	a. (k=2)	b. (k=2)	c. (k=2)
70.3-03.12	0.523	0.549	0.564	0.329	0.743	0.367	**0.318**	0.602	0.357
00.2-10.12	0.902	**0.793**	0.843	0.715*	2.027	0.730*	0.727*	1.423*	0.739*
70.3-10.12	0.481	0.571	0.590	0.377	0.713	0.412	**0.372**	0.866	0.406

Source: Authors´ calculations and Bai and Ng (2008) for PC (10) in the first subsample. The table shows the ratio of MSE of PC, PLS and SPLS over the benchmark model for h=1, 6, 12 and 24. PLS and SPLS results are shown only for the number of components k that yields the best forecasting results. An asterisk (*) means k=1. Bold figures indicate the best forecasting method for each subsample.

Table 3.15
Average number of selected variables for CPI

Period	PC (10)	SPLS											
		h=1			h=6			h=12			h=24		
		a. (k=1)	b. (k=1)	c. (k=1)	a. (k=2)	b. (k=1)	c. (k=2)	a. (k=2)	b. (k=2)	c. (k=2)	a. (k=2)	b. (k=2)	c. (k=2)
70.3-03.12	132	4.886	1.000	80.430	30.738	1.000	9.798	94.317	7.406	83.574	28.571	23.448	28.874
00.2-10.12	112	44.308	28.677	42.631	24.064	1.736	36.144	28.126	1.370	29.303	3.682	15.383	67.206
70.3-10.12	112	44.313	1.000	47.303	91.004	1.190	38.550	37.082	6.912	57.893	28.483	22.998	28.322

Source: Authors´ calculations. The table shows the average number of variables selected by each method for h=1, 6, 12 and 24.

Appendix A. Data definitions and transformations (Stock and Watson 2005)

Short name	Transformation	Mnemonic	Description
PI	Δln	DLPI	Personal income (AR, bil. chain 2000 $)
PI less transfers	Δln	DLPILTRANSFERS	Personal income less transfer payments (AR, bil. chain 2000 $)
Consumption	Δln	DLCONS	Real Consumption (AC) A0m224/gmdc
M&T sales	Δln	DLMTSALES	Manufacturing and trade sales (mil. Chain 1996 $)
Retail sales	Δln	DLRETAILSALES	Sales of retail stores (mil. Chain 2000 $)
IP: total	Δln	DLIPTOTAL	INDUSTRIAL PRODUCTION INDEX - TOTAL INDEX
IP: products	Δln	DLIPPRODUCTS	INDUSTRIAL PRODUCTION INDEX - PRODUCTS, TOTAL
IP: final prod	Δln	DLIPFINALPROD	INDUSTRIAL PRODUCTION INDEX - FINAL PRODUCTS
IP: cons gds	Δln	DLIPCONSGDS	INDUSTRIAL PRODUCTION INDEX - CONSUMER GOODS
IP: cons dble	Δln	DLIPCONSDBLE	INDUSTRIAL PRODUCTION INDEX - DURABLE CONSUMER GOODS
IP: cons nondble	Δln	DLIPCONSNONDBLE	INDUSTRIAL PRODUCTION INDEX - NONDURABLE CONSUMER GOODS
IP: bus eqpt	Δln	DLIPBUSEQPT	INDUSTRIAL PRODUCTION INDEX - BUSINESS EQUIPMENT
IP: materials	Δln	DLIPMATLS	INDUSTRIAL PRODUCTION INDEX - MATERIALS
IP: dble matls	Δln	DLIPDBLEMATLS	INDUSTRIAL PRODUCTION INDEX - DURABLE GOODS MATERIALS
IP: nondble matls	Δln	DLIPNONDBLEMATLS	INDUSTRIAL PRODUCTION INDEX - NONDURABLE GOODS MATERIALS
IP: mfg	Δln	DLIPMFG	INDUSTRIAL PRODUCTION INDEX - MANUFACTURING (SIC)
IP: res util	Δln	DLIPRESUTIL	INDUSTRIAL PRODUCTION INDEX - RESIDENTIAL UTILITIES
IP: fuels	Δln	DLIPFUELS	INDUSTRIAL PRODUCTION INDEX - FUELS
NAPM prodn	lv	NAPMPRODN	NAPM PRODUCTION INDEX (PERCENT)
Cap util	Δlv	DCAPUTIL	Capacity Utilization (Mfg)
Help wanted indx	Δlv	DHELPWANTDIND	INDEX OF HELP-WANTED ADVERTISING IN NEWSPAPERS (1967=100;SA)
Help wanted/emp	Δlv	DHELPWANTEMP	EMPLOYMENT: RATIO; HELP-WANTED ADS:NO. UNEMPLOYED CLF
Emp CPS total	Δln	DLEMPCPSTOTAL	CIVILIAN LABOR FORCE: EMPLOYED, TOTAL (THOUS.,SA)
Emp CPS nonag	Δln	DLEMPCPSNONAG	CIVILIAN LABOR FORCE: EMPLOYED, NONAGRIC.INDUSTRIES (THOUS.,SA)
U: all	Δlv	DUNEMPALL	UNEMPLOYMENT RATE: ALL WORKERS, 16 YEARS & OVER (%,SA)
U: mean duration	Δlv	DUNMEANDUR	UNEMPLOY.BY DURATION: AVERAGE(MEAN)DURATION IN WEEKS (SA)
U< 5 wks	Δln	DLUNL5WKS	UNEMPLOY.BY DURATION: PERSONS UNEMPL.LESS THAN 5 WKS (THOUS.,SA)
U 5-14 wks	Δln	DLUN514WKS	UNEMPLOY.BY DURATION: PERSONS UNEMPL.5 TO 14 WKS (THOUS.,SA)
U 15+ wks	Δln	DLUN15MWKS	UNEMPLOY.BY DURATION: PERSONS UNEMPL.15 WKS + (THOUS.,SA)
U 15-26 wks	Δln	DLUN1526WKS	UNEMPLOY.BY DURATION: PERSONS UNEMPL.15 TO 26 WKS (THOUS.,SA)
U 27+ wks	Δln	DLUN27MWKS	UNEMPLOY.BY DURATION: PERSONS UNEMPL.27 WKS + (THOUS,SA)
UI claims	Δln	DLUICLAIMS	Average weekly initial claims, unemploy. insurance (thous.)
Emp: total	Δln	DLEMPTOTAL	EMPLOYEES ON NONFARM PAYROLLS - TOTAL PRIVATE
Emp: gds prod	Δln	DLEMPGDSPROD	EMPLOYEES ON NONFARM PAYROLLS - GOODS-PRODUCING
Emp: mining	Δln	DLEMPMINING	EMPLOYEES ON NONFARM PAYROLLS - MINING
Emp: const	Δln	DLEMPCONST	EMPLOYEES ON NONFARM PAYROLLS - CONSTRUCTION
Emp: mfg	Δln	DLEMPMFG	EMPLOYEES ON NONFARM PAYROLLS - MANUFACTURING
Emp: dble gds	Δln	DLEMPDBLEGDS	EMPLOYEES ON NONFARM PAYROLLS - DURABLE GOODS
Emp: nondbles	Δln	DLEMPNONDBLES	EMPLOYEES ON NONFARM PAYROLLS - NONDURABLE GOODS
Emp: services	Δln	DLEMPSERV	EMPLOYEES ON NONFARM PAYROLLS - SERVICE-PROVIDING
Emp: TTU	Δln	DLEMPTTU	EMPLOYEES ON NONFARM PAYROLLS - TRADE, TRANSPORTATION, AND UTILITIES
Emp: wholesale	Δln	DLEMPWHSALE	EMPLOYEES ON NONFARM PAYROLLS - WHOLESALE TRADE
Emp: retail	Δln	DLEMPRETAIL	EMPLOYEES ON NONFARM PAYROLLS - RETAIL TRADE
Emp: FIRE	Δln	DLEMPFIRE	EMPLOYEES ON NONFARM PAYROLLS - FINANCIAL ACTIVITIES
Emp: Govt	Δln	DLEMPGOV	EMPLOYEES ON NONFARM PAYROLLS - GOVERNMENT
Emp-hrs nonag	Δln	DLEMPHNONAG	Employee hours in nonag. establishments (AR, bil. hours)
Avg hrs	lv	AVGHRS	AVERAGE WEEKLY HOURS OF PRODUCTION OR NONSUPERVISORY WORKERS ON PRIVATE NONFAR
Overtime: mfg	Δlv	DOVERTMFG	AVERAGE WEEKLY HOURS OF PRODUCTION OR NONSUPERVISORY WORKERS ON PRIVATE NONFAR
Avg hrs: mfg	lv	AVGHMFG	Average weekly hours, mfg. (hours)
NAPM empl	lv	NAPMEMP	NAPM EMPLOYMENT INDEX (PERCENT)
Starts: nonfarm	ln	LSTSNONFARM	HOUSING STARTS:NONFARM(1947-58);TOTAL FARM&NONFARM(1959-)(THOUS.,SA
Starts: NE	ln	LSTSNE	HOUSING STARTS:NORTHEAST (THOUS.U.)S.A.
Starts: MW	ln	LSTSMW	HOUSING STARTS:MIDWEST(THOUS.U.)S.A.
Starts: South	ln	LSTSSOUTH	HOUSING STARTS:SOUTH (THOUS.U.)S.A.
Starts: West	ln	LSTSWEST	HOUSING STARTS:WEST (THOUS.U.)S.A.
BP: total	ln	LBPTOTAL	HOUSING AUTHORIZED: TOTAL NEW PRIV HOUSING UNITS (THOUS.,SAAR)
BP: NE	ln	LBPNE	HOUSES AUTHORIZED BY BUILD. PERMITS:NORTHEAST(THOU.U.)S.A
BP: MW	ln	LBPMW	HOUSES AUTHORIZED BY BUILD. PERMITS:MIDWEST(THOU.U.)S.A.
BP: South	ln	LBPSOUTH	HOUSES AUTHORIZED BY BUILD. PERMITS:SOUTH(THOU.U.)S.A.
BP: West	ln	LBPWEST	HOUSES AUTHORIZED BY BUILD. PERMITS:WEST(THOU.U.)S.A.
PMI	lv	PMI	PURCHASING MANAGERS' INDEX (SA)
NAPM new ordrs	lv	NAPMNWORD	NAPM NEW ORDERS INDEX (PERCENT)
NAPM vendor del	lv	NAPMVDEL	NAPM VENDOR DELIVERIES INDEX (PERCENT)
NAPM Invent	lv	NAPMINVT	NAPM INVENTORIES INDEX (PERCENT)

Short name	Transformation	Mnemonic	Description
Orders: cons gds	Δln	DLORDRCONGDS	Mfrs' new orders, consumer goods and materials (bil. chain 1982 $)
Orders: dble gds	Δln	DLORDRDBLGDS	Mfrs' new orders, durable goods industries (bil. chain 2000 $)
Orders: cap gds	Δln	DLORDRCAPGDS	Mfrs' new orders, nondefense capital goods (mil. chain 1982 $)
Unf orders: dble	Δln	DLUNORDDBLE	Mfrs' unfilled orders, durable goods indus. (bil. chain 2000 $)
M&T invent	Δln	DLMTINVENT	Manufacturing and trade inventories (bil. chain 2000 $)
M&T invent/sales	Δlv	DMTINVTSAL	Ratio, mfg. and trade inventories to sales (based on chain 2000 $)
M1	Δ2ln	DL2M1	MONEY STOCK: M1(CURR,TRAV.CKS,DEM DEP,OTHER CK'ABLE DEP)(BIL$,SA)
M2	Δ2ln	D2LM2	MONEY STOCK:M2(M1+O'NITE RPS,EURO$,G/P&B/D MMMFS&SAV&SM TIME DEP(BIL$,
M3	Δ2ln	DL2M3	MONEY STOCK: M3(M2+LG TIME DEP,TERM RP'S&INST ONLY MMMFS)(BIL$,SA)
M2(real)	Δln	DLM2REAL	MONEY SUPPLY - M2 IN 1996 DOLLARS (BCI)
MB	Δ2ln	DL2MB	MONETARY BASE, ADJ FOR RESERVE REQUIREMENT CHANGES(MIL$,SA)
Reserves tot	Δ2ln	DL2RESERVTOT	DEPOSITORY INST RESERVES:TOTAL,ADJ FOR RESERVE REQ CHGS(MIL$,SA)
Reserves nonbor	Δ2ln	DL2RESERVNONBOR	DEPOSITORY INST RESERVES:NONBORROWED,ADJ RES REQ CHGS(MIL$,SA)
C&I loans	Δ2ln	DL2CILOANS	COMMERCIAL & INDUSTRIAL LOANS OUSTANDING IN 1996 DOLLARS (BCI)
D C&I loans	lv	DELTACILOANS	WKLY RP LG COM'L BANKS:NET CHANGE COM'L & INDUS LOANS(BIL$,SAAR)
Cons credit	Δ2ln	DL2CONSCREDIT	CONSUMER CREDIT OUTSTANDING - NONREVOLVING(G19)
Inst credit/ PI	Δlv	DINSTCREDPI	Ratio, consumer installment credit to personal income (pct.)
S&P 500	Δln	DLSP500	S&P'S COMMON STOCK PRICE INDEX: COMPOSITE (1941-43=10)
S&P: indust	Δln	DLSPINDUST	S&P'S COMMON STOCK PRICE INDEX: INDUSTRIALS (1941-43=10)
S&P div yield	Δlv	DSPDIVYIELD	S&P'S COMPOSITE COMMON STOCK: DIVIDEND YIELD (% PER ANNUM)
S&P PE ratio	Δln	DLSPPERATIO	S&P'S COMPOSITE COMMON STOCK: PRICE-EARNINGS RATIO (%,NSA)
Fed Funds	Δlv	DFEDFUNDS	INTEREST RATE: FEDERAL FUNDS (EFFECTIVE) (% PER ANNUM,NSA)
Comm paper	Δlv	DCOMPAPER	Cmmercial Paper Rate (AC)
3 mo T-bill	Δlv	DTBILL3M	INTEREST RATE: U.S.TREASURY BILLS,SEC MKT,3-MO.(% PER ANN,NSA)
6 mo T-bill	Δlv	DTBILL6M	INTEREST RATE: U.S.TREASURY BILLS,SEC MKT,6-MO.(% PER ANN,NSA)
1 yr T-bond	Δlv	DTBOND1Y	INTEREST RATE: U.S.TREASURY CONST MATURITIES,1-YR.(% PER ANN,NSA)
5 yr T-bond	Δlv	DTBOND5Y	INTEREST RATE: U.S.TREASURY CONST MATURITIES,5-YR.(% PER ANN,NSA)
10 yr T-bond	Δlv	DTBOND10Y	INTEREST RATE: U.S.TREASURY CONST MATURITIES,10-YR.(% PER ANN,NSA)
Aaa bond	Δlv	DAAABOND	BOND YIELD: MOODY'S AAA CORPORATE (% PER ANNUM)
Baa bond	Δlv	DBAABOND	BOND YIELD: MOODY'S BAA CORPORATE (% PER ANNUM)
CP-FF spread	lv	CPFFSPREAD	cp90-fyff
3 mo-FF spread	lv	FFSPREAD3M	fygm3-fyff
6 mo-FF spread	lv	FFSPREAD6M	fygm6-fyff
1 yr-FF spread	lv	FFSPREAD1Y	fygt1-fyff
5 yr-FF spread	lv	FFSPREAD5Y	fygt5-fyff
10 yr-FF spread	lv	FFSPREAD10Y	fygt10-fyff
Aaa-FF spread	lv	AAAFFSPREAD	fyaaac-fyff
Baa-FF spread	lv	BAAFFSPREAD	fybaac-fyff
Ex rate: avg	Δln	DLEXRATEAVG	UNITED STATES;EFFECTIVE EXCHANGE RATE(MERM)(INDEX NO.)
Ex rate: Switz	Δln	DLEXRATESWITZ	FOREIGN EXCHANGE RATE: SWITZERLAND (SWISS FRANC PER U.S.$)
Ex rate: Japan	Δln	DLEXRATEJAPAN	FOREIGN EXCHANGE RATE: JAPAN (YEN PER U.S.$)
Ex rate: UK	Δln	DLEXRATEUK	FOREIGN EXCHANGE RATE: UNITED KINGDOM (CENTS PER POUND)
Ex rate: Canada	Δln	DLEXRATECANADA	FOREIGN EXCHANGE RATE: CANADA (CANADIAN $ PER U.S.$)
PPI: fin gds	Δ2ln	DL2PPIFINGDS	PRODUCER PRICE INDEX: FINISHED GOODS (82=100,SA)
PPI: cons gds	Δ2ln	DL2PPICONSGDS	PRODUCER PRICE INDEX:FINISHED CONSUMER GOODS (82=100,SA)
PPI: int mat'ls	Δ2ln	DL2PPIINTMATLS	PRODUCER PRICE INDEX:INTERMED MAT.SUPPLIES & COMPONENTS(82=100,SA)
PPI: crude mat'ls	Δ2ln	DL2PPICRUDEMAT	PRODUCER PRICE INDEX:CRUDE MATERIALS (82=100,SA)
Spot market price	Δ2ln	DL2SPOTMKPRICE	SPOT MARKET PRICE INDEX:BLS & CRB: ALL COMMODITIES(1967=100)
Sens mat'ls price	Δ2ln	DL2SENSIMATPRICES	INDEX OF SENSITIVE MATERIALS PRICES (1990=100)(BCI-99A)
NAPM com price	lv	NAPMCOMPRICE	NAPM COMMODITY PRICES INDEX (PERCENT)
CPI-U: all	Δ2ln	DL2PUNEW	CPI-U: ALL ITEMS (82-84=100,SA)
CPI-U: apparel	Δ2ln	DL2CPIUAPPAREL	CPI-U: APPAREL & UPKEEP (82-84=100,SA)
CPI-U: transp	Δ2ln	DL2CPIUTRANSP	CPI-U: TRANSPORTATION (82-84=100,SA)
CPI-U: medical	Δ2ln	DL2CPIUMEDICAL	CPI-U: MEDICAL CARE (82-84=100,SA)
CPI-U: comm.	Δ2ln	DL2CPIUCOMM	CPI-U: COMMODITIES (82-84=100,SA)
CPI-U: dbles	Δ2ln	DL2CPIUDBLES	CPI-U: DURABLES (82-84=100,SA)
CPI-U: services	Δ2ln	DL2CPIUSERVICES	CPI-U: SERVICES (82-84=100,SA)
CPI-U: ex food	Δ2ln	DL2CPIUEXFOOD	CPI-U: ALL ITEMS LESS FOOD (82-84=100,SA)
CPI-U: ex shelter	Δ2ln	DL2CPIUEXSHEL	CPI-U: ALL ITEMS LESS SHELTER (82-84=100,SA)
CPI-U: ex med	Δ2ln	DL2CPIUXMED	CPI-U: ALL ITEMS LESS MEDICAL CARE (82-84=100,SA)
PCE defl	Δ2ln	DL2PCEDEFL	PCE,IMPL PR DEFL:PCE (1987=100)
PCE defl: dbles	Δ2ln	DL2PCEDEFLDUR	PCE,IMPL PR DEFL:PCE; DURABLES (1987=100)
PCE defl: nondble	Δ2ln	DL2PCEDEFNONDUR	PCE,IMPL PR DEFL:PCE; NONDURABLES (1996=100)
PCE defl: service	Δ2ln	DL2PCEDEFSERVICE	PCE,IMPL PR DEFL:PCE; SERVICES (1987=100)
AHE: goods	Δ2ln	DL2AHEGOODS	AVERAGE HOURLY EARNINGS OF PRODUCTION OR NONSUPERVISORY WORKERS ON PRIVATE NO
AHE: const	Δ2ln	DL2AHECONST	AVERAGE HOURLY EARNINGS OF PRODUCTION OR NONSUPERVISORY WORKERS ON PRIVATE NO
AHE: mfg	Δ2ln	DL2AHEMFG	AVERAGE HOURLY EARNINGS OF PRODUCTION OR NONSUPERVISORY WORKERS ON PRIVATE NO
Consumer expect	Δlv	DCONSEXP	U. OF MICH. INDEX OF CONSUMER EXPECTATIONS(BCD-83)

In the transformation column, ln denotes logarithm, Δln y Δ2ln denote the first and second difference of the logarithm and lv means level.

Chapter 4

Multivariate Sparse PLS for Macroeconomic Forecasting

4.1 Introduction

In this chapter we extend the ideas and results of the previous one to the problem of forecasting a set of variables simultaneously. The prediction of multiple responses (Y) from a large dataset of predictors (X) is performed regularly in macroeconomics and other fields of study. How to get the best forecast for a given set of targets is one of the most common problems that researchers must face. For example, Stock and Watson (2012) performed a forecast of 143 quarterly U.S. macroeconomic time series, one at a time as dependent variable, as a function of the first five principal components computed from the 109 disaggregate series from the dataset. The remaining 34 series are level aggregates that according to the authors did not add information to the estimated components. In this case, the same factors estimated from the dataset (X) were used to forecast all the responses. Note that these PC factors have good explanatory power for X but may not be good predictors for the targets. There are other approaches that can be implemented in order to improve the accuracy of macroeconomic forecasts. PLS provides an alternative way to cope with this problem, because it explicitly considers the response variable in estimating the factors. In PLS, the factors extracted for forecasting each individual target variable will be specific ones. Therefore, the first estimated components will be the more relevant for predicting a given target. Fuentes, et al. (2014) found for five U.S macroeconomic variables of the Stock and Watson (2005) dataset that univariate PLS, with one or two factors, outperforms the PC method with ten components, for all the subsamples and horizons considered.

Besides prediction of univariate responses, PLS can also work with several response variables. In this instance, the factors are formed taking into account all the considered targets simultaneously and thus, the prediction of each response variable will be based on a common set of jointly estimated factors. In general, the number of targets is much smaller than the dataset of predictors. This feature can be advantageous for forecasting related variables. Furthermore, the property of handling multivariate response is inherited by the SPLS method, and then regularization can also be added to this approach. In the biological context, multivariate PLS has been frequently used showing good prediction results and some forms of regularized multivariate PLS, that have been introduced recently, also seem to achieve high predictive power (Palermo et al., 2009; Chun and Keles, 2010; among others).

Other methods consider the dynamic relationships between two or more variables and perform simultaneous forecasts of the targets. For example, the standard vector autoregression (VAR) model and some proposed extensions that combine the VAR with factor models (FAVAR) in order to incorporate the available information of a large number of predictors (Bernake, et al. 2005; Moench, 2008; Pesaran et al., 2011). Another example is the small scale factor model, such as the proposed by Camacho and Perez-Quiros (2010), which allows to compute forecasts for a set of indicators, although its primary focus is to forecast a specific single target (the euro area GDP growth). As well, other extensions developed for dynamic factor models in order to establish the relevance of adding estimated factors from certain groups of variables (for instance to capture the economic conditions) to improve the forecast accuracy of a different set of variables as, for instance, the yield curves (Koopman and van der Wel, 2013).

In this chapter, we are interested in extending the univariante implementation of the PLS and SPLS methods to the multivariate case. Taking into account some results of the previous literature, which suggest that the inclusion of additional information within a VAR model is helpful to improve the forecast accuracy of economic and financial variables, we adopt the VAR framework as a natural extension for the multivariate case. First, we reappraise the question of how to add potential relevant information into a multivariate forecasting framework. Considering the factor based models, we explore if the extraction of jointly estimated factors for a specific group of response variables, instead of the construction of generic common factors, such as PC factors, may improve the forecast performance of the targets. Second, within this context we investigate if discarding some variables with low information content may help to further improve the forecasts accuracy. Third, with the purpose of forecasting, we examine within the PLS framework, if the dynamics of a series could be better capture only through its own lags (in a parsimonious way) instead of considering lags of all the target variables.

We use the Stock and Watson database in order to compare the forecasting performance of multivariate PLS and SPLS techniques and other multivariate approaches such as the standard VAR and the VAR augmented with factors (VARF). Notice that this last model differs from the FAVAR model because the estimated factors are not incorporated as endogenous variables but as exogenous ones. In this sense, the VARF model is closer to the VARX model.

We employ a set of four target variables that provide information about the current state of the economy: industrial production (IP), total employment (EMT), personal income (PI) and retail sales (RS).

The results show that the models that incorporate factors oriented to the targets give better forecasting results than alternative models with no oriented factors. Additionally, the sparse version yields further improvements in the forecasting results. The degree of refinement increases with the forecast horizon.

This chapter is organized as follows. Section 4.2 describes the forecasting framework and some particular features of the VAR enlarged with factors, the PLS and SPLS multivariate methods. Section 4.3 reports the comparative results of the empirical application. Section 4.4 concludes.

4.2 Forecasting Framework

Our goal is to predict a small set of target variables as a function of their own lags, the lags of the other target variables and a large dataset of predictors (X_t). In order to perform the prediction we will implement the traditional process in large dynamic factor models that consists of two steps (see, for instance, Stock and Watson, 2002). In the first step, the data dimensionality is reduced by means of estimating the common factors and in a second step, once the factors have been estimated, they are used as predictors in the forecasting framework. We use a multivariate forecasting framework. In particular, it corresponds to the VARF model

$$Y_{t+h} = C + \Phi(L)Y_t + \Lambda\hat{F}_t + \mathrm{a}_{t+h} \tag{4.1}$$

where Y_{t+h} is the set of n target variables to be forecasted, n << N, N is the number of predictors in X_t, C is a nx1 vector of constants, Φ(L) is a conformable lag polynomial matrix that relates the target with its past values and those of the other variables in the VAR, Λ is an nxr matrix of parameters that measures the relation between the variables to be forecasted and the current realizations of the estimated factors ($\hat{F}$) and $a_{t+h,}$ is the vector of h step ahead forecast errors terms with zero mean and covariance matrix Σ.

Consider, for example, the VARF(1) model for three variables and one estimated factor, f_t^1:

$$\begin{bmatrix} y_{1,t+h} \\ y_{2,t+h} \\ y_{3,t+h} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} + \begin{bmatrix} \phi_{11} & \phi_{12} & \phi_{13} \\ \phi_{21} & \phi_{22} & \phi_{23} \\ \phi_{31} & \phi_{32} & \phi_{33} \end{bmatrix} \begin{bmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \end{bmatrix} + \begin{bmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \end{bmatrix} \begin{bmatrix} f_t^1 \end{bmatrix} + \begin{bmatrix} a_{1,t+h} \\ a_{2,t+h} \\ a_{3,t+h} \end{bmatrix} \tag{4.2}$$

In the multivariate case, we consider the entire set of responses when extracting the multivariate PLS and SPLS factors from X_t. Then, a common set of factors is estimated for forecasting the n target variables in the considered set (Y_t).

We estimate the forecasting models including different types of restrictions: (i) to consider the dynamic relationship between the observable variables, we compute a standard VAR model without factors, (ii) to take into account the importance of incorporating additional information we enlarge the VAR model first, with PC factors and later, in order to examine whether the way in which the factors are extracted is relevant for the forecasting performance, we augmented the VAR model with PLS and SPLS factors; finally, (iii) to explore if the way in which the dynamics of a time series is incorporated to the model (univariate or multivariate) it is relevant for the forecasting performance, we compute a VAR model where the matrices Σ and Φ(L) are diagonal and it is augmented with PLS and SPLS factors.

All the models that include a VAR structure are estimated with four lags and, alternatively, selecting the optimal number of lags of the autoregressive order and the factor order by the Bayesian Information Criterion (BIC).

4.2.1 Partial Least Squares (PLS)

The PLS method can handle univariate and multivariate responses. When the problem incorporates multivariate targets is referred as *PLS2*. Although the objective function of PLS2 can be defined in the same way it was done for univariate PLS, the formulation criterion to find the direction vectors is different. In the univariate case, PLS seeks directions that have high variance and have high correlation with the response (Hastie et al., 2008):

$$\begin{aligned} &w_i = \arg\max_w \operatorname{cov}(y, Xw) \equiv \arg\max_w corr^2(y, Xw)\ \operatorname{var}(y)\ \operatorname{var}(Xw) \\ &subject\ to\ w'w = 1. \end{aligned} \tag{4.3}$$

Since var(y) does not depend on w, this term is excluded from the objective function:

$$\begin{aligned} &w_i = \arg\max_w corr^2(y, Xw) \operatorname{var}(Xw) \\ &subject\ to\ w'w = 1; \end{aligned} \tag{4.4}$$

whereas in the multivariate case, PLS2 seeks to find linear combinations of the predictors (t=Xw) and linear combinations of the responses (u=Yv) that maximize the covariance between them:

$$\begin{aligned} &w_i = \arg\max_w \operatorname{cov}^2(Yv, Xw) \equiv \arg\max_w \operatorname{var}(Yv)\ corr^2(Yv, Xw)\ \operatorname{var}(Xw) \\ &subject\ to\ v'v = w'w = 1. \end{aligned} \tag{4.5}$$

The dimension reduction is performed in both matrices X and Y. Here, the variances of the response matrix as well as the predictor matrix are involved. The estimated components should try to explain most variation in all the responses simultaneously as well as the predictor's variance.

Rosipal and Krämer (2006) examine the connection between PLS and Canonical Correlation Analysis (CCA) by means of their objective functions. They pointed out that the criterion of CCA of finding the maximal correlation between X and Y: $w_i = \arg\max_w corr^2(Yv, Xw)$, is balanced in PLS when adding the condition to explain as much variance as possible in X and Y. Additionally, Sun et al. (2011) highlight that CCA reduces to a generalized eigenvalue problem, while the classical solution for the multivariate PLS2 problem is based on an iterative process. The estimated components are obtained when the sequence of repeated estimates of the vectors w and v converges. For the next components, the X and Y matrices are subsequently deflated with respect to the preceding component or factor, where the residuals matrices required for the successive component estimation process are obtained by regressing each target against the previous component and then, forming a linear combination of them.

De Jong (1993) proposed another way to solve the univariante and multivariate PLS problem, the SIMPLS algorithm in which the weighting vectors are estimated directly from the original matrices.

According to Wold et al. (2001, 2004), PLS2 might be useful when the targets are related to or measure the same concept. Palermo et al. (2009) highlight that multivariate PLS will allow exploiting relationships between responses. It is important to note that the same resulting PLS2 components will be used in the forecasting equation of each target variable, even though it is expected that the regression coefficients will be different.

We examine the performance of the **static approach** for the PLS2 in a time series context. In the previous chapter, we consider different ways to apply the PLS methods and we found that this approach is able to generate good predictive results. In this case, the components are extracted by applying PLS between the jointly set of target variables and the original set of predictors (X). Then, the lags of the target variables are included in the forecasting equations (4.1).

Later, we examine a more parsimonious model where we consider two of the univariate proposed approaches also for PLS2: the static one and the ***dynamic approach.*** In the first case, the $\Phi(L)$ matrix in (4.1) is restricted to be diagonal. The dynamic approach consists in extracting the components based on applying PLS2 over the residual matrix from the AR(p) process fitted for each target variable. In this case, we believe that the dynamics of each time series are isolated in a proper way and then, it is valid construct PLS2 factors from the matrix of residuals. By the contrary, if we apply a VAR(q) model instead of an AR(p) process, we eliminate the potential correlations between the target variables and then, it does not make sense to apply PLS2 to the residual matrix of this process. As in the static case, the lags of each target variable are included in the forecasting equations (4.1).

As it was mentioned before, the SPLS method also can be implemented in the multivariate case (SPLS2). In this instance, the optimization problem given in (3.13) is solving through iterations between the solutions for the original vector (w) for a fixed surrogate direction vector (c) and alternatively, the solutions for the surrogate direction vector (c) for a fixed original vector (w). For univariate SPLS the solution for the first component does not require this iterative process (Chun and Keles, 2010). When there are $n > 1$ targets, a second penalty L2 becomes relevant in the solution of the objective function (3.14). This penalty takes care of potential singularity in M when solving for c, and can be handled by the parameter θ.

We apply the previous approaches of PLS for SPLS, in the multivariate case.

4.3 Empirical Application

In order to compare the performance of the multivariate PLS and SPLS with the univariate PLS and SPLS models, we employ the Stock and Watson (2005) database. The data consists of 132 major monthly macroeconomic variables from United States (U.S.) over the period 1960:1-2003:12, hence N = 132 and T = 528. As was mentioned before, the target set is formed by the following indicators: industrial production (IP), total employment (EMT), personal income (PI) and retail sales (RS). These variables are pointed out by the NBER as the four monthly series that should be checked jointly with quarterly gross domestic product (GDP) in the definition of a recession. Some of them are frequently considered as coincident indicators and thus, they are included in the construction of the corresponding business cycle indicator (see for instance, The Conference Board, Federal Reserve Bank of St. Louis, New York and Dallas). We assume that the log level of the four series is integrated of order 1 and are defined as:

$$y_{t+h}^{h} = \frac{1200}{\mathrm{h}}(y_{t+h} - y_t). \tag{4.6}$$

Let

$$z_t = 1200(y_t - y_{t-1}). \tag{4.7}$$

With the aim of facilitating the comparison exercise, we consider seven subsamples, as defined by Bai and Ng (2008) and Fuentes et al. (2014). Table 4.1 summarizes the estimation and forecast subsamples. Additionally, we perform the estimations for horizons of 1, 6, 12 and 24 months.

For the multivariate case the factors are estimated jointly for the group of variables. The target variables are removed from the set of predictors for the factor estimation.

4.3.1 Forecast Results

To compare the performance of the PLS and SPLS methods we use as benchmark an AR(4). The forecast performance is evaluated using the relative mean-squared forecast errors (RMSE) over the benchmark:

$$RMSE\ (method) = \frac{MSE\ (method)}{MSE\ (AR(4))}. \tag{4.8}$$

An entry of less than one implies an improvement of the method upon the simple AR(4) forecast.

Tables 4.2 to 4.9 show the forecasting results for the different variables and horizons considered. We estimate a VAR model enlarged with the first ten PC factors and with the first two components extracted with the PLS and SPLS methods. Additionally for SPLS models, the estimated factors were computed for four values of the sparsity parameter {0.2, 0.4, 0.6, 0.8}.

VARF Models

a. RMSE

The entries in Tables 4.2 to 4.5 report some interesting findings. First, in general the results point out that both the dynamic relationship between the variables and the inclusion of potential relevant information for forecasting into the multivariate structure appears to be more significant for longer horizons. The results of the bottom panel of the Tables 4.2 to 4.5 highlight that the construction of factors by means of PLS and SPLS improves the forecast efficiency of the VARF methods. In particular, the VARF extended with SPLS2 factors outperforms in 85.7% of the cases the VAR models and the alternative VARF models, which incorporates PC or PLS factors.

For h=1, the VAR(4) model, that constitutes the simplest process considered, it is able to generate small predictive gains compared to the benchmark model, although in some cases its performance is worse. Then, when the VAR(4) model is augmented with PC(10) factors, the RMSE deteriorates. In this case, it seems that the uncertainty due to the increase in the number of estimated parameters is greater than the contribution made by the PC(10) factors. Conversely, when the VAR(4) is enlarged with PLS2 and SPLS2 factors the RMSE shows an improvement with respect to the previous models. In particular, the inclusion of SPLS2 factors in the models yields the best forecasting performance in 64.3% of cases. These results underline the fact that the way in which the factors are constructed is important for the final forecast. Extracting specific jointly factors for a set of given targets and even more, discarding some predictors might produce lower RMSE.

For h=6, with the exception of PI, the AR(4) model outperforms the VAR(4) model for the other three variables. However, for this horizon the inclusion of PC(10) factors performs better than the VAR(4) model, outperforming the benchmark in most of the samples. It’s important to note that in this case the VARF models estimated with an optimal number of autoregressive lags, according to the BIC, provide in general better results. In a similar way that in the previous horizon, the VAR augmented with SPLS2 factors is the best performing model in 92.9% of the cases.

For h=12 and 24, the competitive models outperform the benchmark for almost all the horizons and variables, with a clear exception of the sample 3, where the AR(4) model performs best in the major part of the cases. Similar to what was observed for h=6, in general the VARF with SPLS2 factors provides the best forecast results for h=12 and h=24.

b. The number of selected predictors

The number of selected predictors for the best VARF model with SPLS2 factors differs by horizon and sample. To have an idea about these differences, we observe the results for the sample 7. For this case, the average number of chosen variables for each horizon oscillates between 11.8 and 86.4. The degree of sparsity is low (0.4) for the two first horizons and even lower for h=12 (0.2). On the contrary, h=24 is the horizon with the lowest number of selected variables, accordingly the corresponding model is estimated with one component and the degree of sparsity is the highest (λ=0.6). As regards the variables chosen in this subsample, we observe that for h=1 the employment and production indicators are the most frequent variables selected, while several price indicators and monetary variables are discarded. For h=6, besides the employment and production indicators there are other groups of variables such as the financial spreads and the M2 real aggregate within the most frequently chosen variables. In this case, the weights of monetary variables as well as the price indicators have been set to zero. For h=12, the financial spreads and the real M2 aggregate dominate the list of selected variables. In the other side, the weights of monetary variables, prices and indicators of average hourly earnings of production are zeroed out. Finally, for h=24 the most selected variables are financial spreads and the real M2 aggregate.

Restricted VARF Models

Tables 4.6 to 4.9 summarize the results for the two ways in which the restricted version of the multivariate PLS and SPLS method have been applied. The Tables show that in almost 80% of cases, these models yield better forecasting results than the VAR models enlarged with the SPLS2 factors. Notice that the relative performance of SPLS2 models improves with the forecast horizon, for h=12 the VARF models outperform the restricted versions in 40% of all the samples and variables and for h=24, in 29%.

4.4 Conclusions

The results confirm that incorporating potentially relevant information into a multivariate forecasting framework may yield predictive gains. The VAR models augmented with multivariate PLS2 and SPLS2 factors improve the forecast performance of almost all the subsamples, horizons and variables considered, respect to its competitors. In particular, the construction of jointly specific factors from a subset of relevant predictors, by means of the SPLS2 method, shows a relative better performance.

The performance of the VARF models, that attempt to capture simultaneously the dynamic relationship between the target variables and its relation with other potential informative predictors, improves with the forecast horizon. When the construction of the factors takes into account the set of targets, the VARF model is able to outperform the benchmark and the standard VAR model for shorter horizons (h=1). However, when the generic PC factors are used, the RMSE worsens.

Additionally, within the possibilities of the PLS framework analyzed it seems that the dynamic of a time series is better capture trough its own lags. Hence, the restricted model yields better forecasting performance. That is, including lags of the other targets in the set tends to deteriorate the individual forecasting performance. In this sense a more parsimonious model is advisable.

References

Bernanke, B., Boivin, J. and Eliazs, P. (2005). “Measuring the effects of Monetary Policy: A Factor Augmented Vector Autoregressive (FAVAR) Approach”. The Quarterly Journal of Economics, 120 (1): 387-422.

Bai, J. and S. Ng (2008). “Forecasting Economic Time Series Using Targeted Predictors”. Journal of Econometrics, 146: 304-317.

Camacho, M. and G. Perez-Quiros (2010). “Introducing the Euro-Sting: Short-Term Indicator of Euro Area Growth”. *Journal of Applied Econometrics*, 25: 663-694.

Chun H. and S. Keles (2010). “Sparse Partial Squares Regression for Simultaneous Dimension Reduction and Variable Selection”, Journal of the Royal Statistical Society. Series B, Statistical Methodology, 72(1): 3-25.

De Jong, S. (1993). "Simpls: An Alternative Approach to Partial Least Squares Regression". *Chemometrics Intelligent Laboratory Systems,* 18: 251–263.

Fuentes, J., Poncela, P. and J. Rodríguez (2014). "Sparse Partial Least Squares in Time Series for Macroeconomic Forecasting". *Journal of Applied Econometric,* Forthcoming.

Hastie T., Tibshirani, R. and J. Friedman (2008). *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer Series in Statistics, 2nd. edition.

Koopman, S. J and van der Wel, M. (2013). "Forecasting the U.S. Term Structure of Interest Rates Using a Macroeconomic Smooth Dynamic Factor Model. *International Journal of Forecasting*, 29 (4): 676-694.

Moench, E. (2008). "Forecasting the Yield Curve in a Data-Rich Environment: A No-Arbitrage Factor-Augmented VAR approach", Journal of Econometrics, 146: 26–43.

Palermo, G. 2009, Piriano, P. and H-D Zucht (2009). "Performance of PLS Regression Coefficients in Selecting Variables for each Response of a Multivariate PLS for Omics-Type Data". *Advances and Applications in Bioinformatics and Chemestry*, 2:57-70.

Pesaran, M., Pick, A. and Timmermann, A. (2011). "Variable Selection, Estimation and Inference for Multi-period Forecasting Problems", *Journal of Econometrics*, 164:173-187.

Rosipal, R. and N. Krämer (2006). Subspace, Latent Structure and Feature Selection. "Overview and Recent Advances in Partial Least Squares". *Lecture Notes in Computer Science*, 3940: 34-51.

Stock, J. and M. Watson (2002). "Macroeconomic Forecasting Using Diffusion Indexes". *Journal of Business & Economic Statistics*, 20: 147-163.

Stock, J. and M. Watson (2005). "Implications of Dynamic Factor Models for VAR Analysis". NBER Working Paper No. 11467.

Stock, J. and M. Watson (2012). "Generalized Shrinkage Methods for Forecasting Using Many Predictors". *Journal of Business & Economic Statistics*, 30(4): 481-493.

Sun, L., Ji, S. and Y. Jieping (2011). "Canonical Correlation Analysis for Multilabel Classification: A Least Square Formulation, Extensions and Analysis". *Pattern Analysis and Machine Intelligence, IEEE Transactions,* 33(1):194-200.

Wold S., Sj*öström*, M. and L. Ericksson (2001). "PLS-Regression: A Basic Tool of Chemometrics", Chemometrics and Intelligent Laboratory Systems, 58:109-130.

Wold, S., Eriksson, L.,Trygg, J. and N. Kettaneh (2004). "The PLS method-partial least squares projections to latent structures and its applications in industrial RDP". Presented at the COMPSTAT 2004, 16th Symposium of IASC, Prague.

Table 4.1
Estimation and forecast subsamples

SS	Estimation subsample	Forecast subsample
M1	1960:03 to 1970:03-h	1970:03 to 1980:12
M2	1960:03 to 1980:03-h	1980:03 to 1990:12
M3	1960:03 to 1990:03-h	1990:03 to 2000:12
M4	1960:03 to 1970:03-h	1970:03 to1990:12
M5	1960:03 to 1970:03-h	1970:03 to 2000:12
M6	1960:03 to 1980:03-h	1980:03 to 2000:12
M7	1960:03 to 1970:03-h	1970:03 to 2003:12

Table 4.2. RMSE, h=1

Industrial Production

Period	VAR(4)	VARF					
		VAR(4) + PC	VAR(BIC) + PC	VAR(4) + PLS2	VAR(BIC) + PLS2	VAR(4) + SPLS2	VAR(BIC) + SPLS2
70.3-80.12	0.980	1.160	1.101	0.914	0.879	**0.857**	0.875
80.3-90.12	0.968	1.192	1.173	0.748	0.798	0.724	**0.711**
90.3-00.12	**0.969**	1.010	1.026	1.019	1.033	1.011	1.073
70.3-90.12	0.990	1.147	1.087	0.889	0.873	**0.866**	0.899
70.3-00.12	0.990	1.122	1.068	0.904	0.900	**0.877**	0.932
80.3-00.12	0.973	1.133	1.122	0.843	0.845	0.843	**0.842**
70.3-03.12	0.989	1.122	1.068	0.929	0.903	0.909	**0.899**

Total Employment

Period	VAR(4)	VARF					
		VAR(4) + PC	VAR(BIC) + PC	VAR(4) + PLS2	VAR(BIC) + PLS2	VAR(4) + SPLS2	VAR(BIC) + SPLS2
70.3-80.12	0.961	1.111	1.077	0.874	0.929	**0.847**	1.073
80.3-90.12	1.027	1.119	1.128	**0.849**	0.920	0.868	0.861
90.3-00.12	1.250	1.266	1.275	1.213	**1.143**	1.228	1.188
70.3-90.12	0.988	1.093	1.057	0.880	0.914	**0.866**	0.947
70.3-00.12	1.008	1.108	1.068	0.898	0.902	**0.878**	0.981
80.3-00.12	1.063	1.147	1.152	**0.921**	0.922	0.923	0.956
70.3-03.12	1.017	1.133	1.108	0.942	0.938	0.934	**0.922**

Personal Income

Period	VAR(4)	VARF					
		VAR(4) + PC	VAR(BIC) + PC	VAR(4) + PLS2	VAR(BIC) + PLS2	VAR(4) + SPLS2	VAR(BIC) + SPLS2
70.3-80.12	0.955	1.077	1.068	0.913	0.927	**0.891**	0.950
80.3-90.12	1.003	1.171	1.135	1.038	1.056	**0.990**	1.122
90.3-00.12	**0.945**	0.973	0.963	0.952	0.949	0.952	0.963
70.3-90.12	0.984	1.101	1.060	0.965	**0.944**	0.945	0.997
70.3-00.12	0.965	1.036	1.017	0.974	**0.944**	0.973	0.990
80.3-00.12	**0.961**	1.226	1.012	0.975	0.982	0.976	1.020
70.3-03.12	0.963	1.313	1.010	0.957	0.946	0.950	**0.930**

Retail Sales

Period	VAR(4)	VARF					
		VAR(4) + PC	VAR(BIC) + PC	VAR(4) + PLS2	VAR(BIC) + PLS2	VAR(4) + SPLS2	VAR(BIC) + SPLS2
70.3-80.12	1.021	1.054	1.076	0.963	1.069	**0.954**	1.044
80.3-90.12	1.037	1.074	1.054	**0.955**	0.968	**0.967**	1.013
90.3-00.12	1.284	1.309	1.295	1.214	1.233	1.212	1.229
70.3-90.12	1.034	1.057	1.017	0.975	0.962	**0.960**	0.969
70.3-00.12	1.082	1.101	1.080	1.042	**1.025**	1.034	1.035
80.3-00.12	1.116	1.143	1.138	1.032	1.060	**1.032**	1.104
70.3-03.12	1.072	1.843	1.064	1.034	1.012	1.018	**0.999**

Average RMSE

Period	VAR(4)	VARF					
		VAR(4) + PC	VAR(BIC) + PC	VAR(4) + PLS2	VAR(BIC) + PLS2	VAR(4) + SPLS2	VAR(BIC) + SPLS2
70.3-80.12	0.979	1.101	1.081	0.916	0.951	**0.887**	0.985
80.3-90.12	1.009	1.139	1.123	0.898	0.935	**0.887**	0.927
90.3-00.12	1.112	1.139	1.140	1.099	**1.089**	1.101	1.113
70.3-90.12	0.999	1.099	1.055	0.927	0.923	**0.909**	0.953
70.3-00.12	1.011	1.092	1.058	0.954	0.943	**0.941**	0.985
80.3-00.12	1.028	1.162	1.106	**0.943**	0.952	0.944	0.981
70.3-03.12	1.010	1.353	1.063	0.966	0.950	0.953	**0.938**

Source: Authors' calculations. The table shows the ratio of RMSE of VAR(4), VAR(4) and VAR(BIC) +PC, +PLS2 and +SPLS2 over the benchmark model for h=1. Bold figures indicate the best forecasting method for each subsample. The bottom panel shows the average RMSE.

Table 4.3. RMSE, h=6

Industrial Production

Period	VAR(4)	VARF					
		VAR(4) + PC	VAR(BIC) + PC	VAR(4) + PLS2	VAR(BIC) + PLS2	VAR(4) + SPLS2	VAR(BIC) + SPLS2
70.3-80.12	1.035	0.645	0.633	0.466	0.463	**0.427**	0.473
80.3-90.12	1.007	0.986	0.985	**0.703**	0.863	0.708	0.740
90.3-00.12	1.064	1.304	1.454	0.893	1.011	0.880	**0.846**
70.3-90.12	1.025	0.771	0.753	0.585	0.568	0.538	**0.532**
70.3-00.12	1.032	0.829	0.833	0.605	0.614	0.596	**0.570**
80.3-00.12	1.025	1.068	1.098	0.752	0.831	0.757	**0.743**
70.3-03.12	1.035	0.877	0.884	0.678	0.686	**0.642**	0.661

Total Employment

Period	VAR(4)	VARF					
		VAR(4) + PC	VAR(BIC) + PC	VAR(4) + PLS2	VAR(BIC) + PLS2	VAR(4) + SPLS2	VAR(BIC) + SPLS2
70.3-80.12	1.017	0.867	0.828	0.560	0.634	**0.558**	0.632
80.3-90.12	1.052	0.949	0.961	0.697	0.883	0.706	**0.691**
90.3-00.12	1.006	1.163	1.176	0.904	0.933	**0.896**	0.903
70.3-90.12	1.011	0.882	0.842	0.615	0.621	0.619	**0.593**
70.3-00.12	1.013	0.913	0.876	0.629	0.648	0.627	**0.626**
80.3-00.12	1.050	1.111	1.017	0.743	0.797	0.750	**0.701**
70.3-03.12	1.019	0.959	0.924	0.711	0.715	0.708	**0.695**

Personal Income

Period	VAR(4)	VARF					
		VAR(4) + PC	VAR(BIC) + PC	VAR(4) + PLS2	VAR(BIC) + PLS2	VAR(4) + SPLS2	VAR(BIC) + SPLS2
70.3-80.12	1.003	0.814	0.769	0.605	0.618	0.586	**0.561**
80.3-90.12	0.915	0.853	0.824	0.806	**0.799**	0.801	0.864
90.3-00.12	0.973	1.157	1.170	1.009	1.057	1.003	**0.946**
70.3-90.12	0.965	0.841	0.807	0.760	0.744	**0.689**	0.733
70.3-00.12	0.967	0.935	0.907	0.785	0.830	**0.770**	0.788
80.3-00.12	0.938	1.620	1.007	0.882	0.974	**0.875**	0.922
70.3-03.12	0.959	0.930	0.902	0.849	0.864	**0.804**	0.848

Retail Sales

Period	VAR(4)	VARF					
		VAR(4) + PC	VAR(BIC) + PC	VAR(4) + PLS2	VAR(BIC) + PLS2	VAR(4) + SPLS2	VAR(BIC) + SPLS2
70.3-80.12	1.021	0.867	0.844	0.564	0.512	**0.505**	0.507
80.3-90.12	1.057	0.777	0.786	0.576	0.622	**0.571**	0.581
90.3-00.12	1.157	1.215	1.159	1.064	1.054	1.058	**1.022**
70.3-90.12	1.035	0.841	0.796	0.623	0.590	0.617	**0.560**
70.3-00.12	1.060	0.922	0.874	0.724	0.681	0.715	**0.658**
80.3-00.12	1.093	0.939	0.912	0.764	0.775	0.762	**0.753**
70.3-03.12	1.062	0.932	0.899	0.716	0.700	0.718	**0.683**

Average RMSE

Period	VAR(4)	VARF					
		VAR(4) + PC	VAR(BIC) + PC	VAR(4) + PLS2	VAR(BIC) + PLS2	VAR(4) + SPLS2	VAR(BIC) + SPLS2
70.3-80.12	1.019	0.798	0.769	0.549	0.557	**0.519**	0.543
80.3-90.12	1.008	0.891	0.889	**0.695**	0.792	0.697	0.719
90.3-00.12	1.050	1.210	1.240	0.967	1.014	0.959	**0.929**
70.3-90.12	1.009	0.834	0.800	0.646	0.631	0.616	**0.605**
70.3-00.12	1.018	0.900	0.873	0.686	0.693	0.677	**0.660**
80.3-00.12	1.027	1.184	1.008	0.785	0.844	0.786	**0.780**
70.3-03.12	1.019	0.924	0.902	0.738	0.741	**0.718**	0.722

Source: Authors' calculations. The table shows the ratio of RMSE of VAR(4), VAR(4) and VAR(BIC) +PC, +PLS2 and +SPLS2 over the benchmark model for h=6. Bold figures indicate the best forecasting method for each subsample. The bottom panel shows the average RMSE.

Table 4.4. RMSE, h=12

Industrial Production

Period	VAR(4)	VARF					
		VAR(4) + PC	VAR(BIC) + PC	VAR(4) + PLS2	VAR(BIC) + PLS2	VAR(4) + SPLS2	VAR(BIC) + SPLS2
70.3-80.12	0.772	0.353	0.375	0.255	0.215	0.206	**0.182**
80.3-90.12	0.982	0.779	0.784	0.626	**0.578**	0.624	0.580
90.3-00.12	1.482	1.649	1.640	1.202	1.207	1.204	**1.159**
70.3-90.12	0.870	0.521	0.565	0.408	0.407	**0.392**	0.397
70.3-00.12	0.926	0.624	0.650	0.476	0.470	0.473	**0.461**
80.3-00.12	1.094	0.975	0.984	0.750	**0.705**	0.756	0.746
70.3-03.12	0.951	0.686	0.717	**0.533**	0.543	0.542	0.546

Total Employment

Period	VAR(4)	VARF					
		VAR(4) + PC	VAR(BIC) + PC	VAR(4) + PLS2	VAR(BIC) + PLS2	VAR(4) + SPLS2	VAR(BIC) + SPLS2
70.3-80.12	0.844	0.618	0.638	0.376	0.357	**0.348**	0.359
80.3-90.12	1.029	0.790	0.789	0.629	0.549	0.628	**0.538**
90.3-00.12	1.114	1.154	1.113	0.952	**0.914**	0.960	0.938
70.3-90.12	0.934	0.706	0.700	0.495	**0.469**	0.508	0.473
70.3-00.12	0.962	0.771	0.764	0.554	0.533	0.562	**0.532**
80.3-00.12	1.067	0.993	0.895	0.725	**0.630**	0.737	0.649
70.3-03.12	0.976	0.822	0.816	0.613	**0.595**	0.630	0.596

Personal Income

Period	VAR(4)	VARF					
		VAR(4) + PC	VAR(BIC) + PC	VAR(4) + PLS2	VAR(BIC) + PLS2	VAR(4) + SPLS2	VAR(BIC) + SPLS2
70.3-80.12	0.852	0.720	0.716	0.591	0.560	0.506	**0.466**
80.3-90.12	0.901	0.811	0.817	0.733	0.733	0.726	**0.718**
90.3-00.12	**1.089**	1.293	1.307	1.105	1.116	1.098	1.102
70.3-90.12	0.886	0.782	0.773	0.691	0.660	0.621	**0.617**
70.3-00.12	0.942	0.927	0.923	0.804	0.784	0.771	**0.748**
80.3-00.12	0.996	1.611	1.073	**0.911**	0.912	0.920	0.929
70.3-03.12	0.947	0.921	0.918	0.823	0.806	**0.799**	0.802

Retail Sales

Period	VAR(4)	VARF					
		VAR(4) + PC	VAR(BIC) + PC	VAR(4) + PLS2	VAR(BIC) + PLS2	VAR(4) + SPLS2	VAR(BIC) + SPLS2
70.3-80.12	0.799	0.646	0.617	0.485	0.472	0.431	**0.421**
80.3-90.12	0.947	0.679	0.691	0.525	**0.511**	0.524	0.514
90.3-00.12	1.287	1.258	1.237	1.125	1.115	1.122	**1.047**
70.3-90.12	0.874	0.677	0.656	0.521	0.497	0.481	**0.479**
70.3-00.12	0.971	0.816	0.793	0.654	0.630	0.621	**0.603**
80.3-00.12	1.097	0.930	0.929	0.772	**0.752**	0.782	0.765
70.3-03.12	0.983	0.826	0.808	0.669	0.648	**0.636**	0.644

Average RMSE

Period	VAR(4)	VARF					
		VAR(4) + PC	VAR(BIC) + PC	VAR(4) + PLS2	VAR(BIC) + PLS2	VAR(4) + SPLS2	VAR(BIC) + SPLS2
70.3-80.12	0.817	0.584	0.586	0.427	0.401	0.373	**0.357**
80.3-90.12	0.965	0.765	0.770	0.628	0.592	0.625	**0.588**
90.3-00.12	1.243	1.339	1.324	1.096	1.088	1.096	**1.061**
70.3-90.12	0.891	0.671	0.674	0.529	0.508	0.500	**0.491**
70.3-00.12	0.950	0.784	0.783	0.622	0.604	0.607	**0.586**
80.3-00.12	1.063	1.127	0.970	0.790	**0.750**	0.799	0.772
70.3-03.12	0.964	0.814	0.815	0.659	0.648	0.652	**0.647**

Source: Authors' calculations. The table shows the ratio of RMSE of VAR(4), VAR(4) and VAR(BIC) +PC, +PLS2 and +SPLS2 over the benchmark model for h=12. Bold figures indicate the best forecasting method for each subsample. The bottom panel shows the average RMSE.

Table 4.5. RMSE, h=24

Industrial Production

Period	VAR(4)	VARF					
		VAR(4) + PC	VAR(BIC) + PC	VAR(4) + PLS2	VAR(BIC) + PLS2	VAR(4) + SPLS2	VAR(BIC) + SPLS2
70.3-80.12	0.723	0.476	0.470	0.433	0.398	0.307	**0.293**
80.3-90.12	0.745	0.463	0.536	0.451	**0.383**	0.389	0.420
90.3-00.12	1.273	1.383	1.376	1.036	1.064	**0.993**	1.009
70.3-90.12	0.760	0.496	0.529	0.450	0.433	**0.341**	0.358
70.3-00.12	0.828	0.602	0.620	0.497	0.498	**0.427**	0.438
80.3-00.12	0.904	0.712	0.764	0.645	**0.579**	0.597	0.601
70.3-03.12	0.847	0.615	0.632	0.518	0.482	0.461	**0.451**

Total Employment

Period	VAR(4)	VARF					
		VAR(4) + PC	VAR(BIC) + PC	VAR(4) + PLS2	VAR(BIC) + PLS2	VAR(4) + SPLS2	VAR(BIC) + SPLS2
70.3-80.12	0.923	0.730	0.683	0.524	0.509	0.432	**0.423**
80.3-90.12	0.864	0.618	0.640	0.483	**0.410**	0.447	0.423
90.3-00.12	1.028	1.076	1.083	0.939	0.949	**0.937**	0.945
70.3-90.12	0.903	0.692	0.669	0.506	0.460	0.444	**0.427**
70.3-00.12	0.928	0.764	0.747	0.572	0.540	0.537	**0.533**
80.3-00.12	0.926	0.783	0.804	0.673	**0.617**	0.617	0.624
70.3-03.12	0.945	0.775	0.758	0.593	0.569	0.571	**0.547**

Personal Income

Period	VAR(4)	VARF					
		VAR(4) + PC	VAR(BIC) + PC	VAR(4) + PLS2	VAR(BIC) + PLS2	VAR(4) + SPLS2	VAR(BIC) + SPLS2
70.3-80.12	0.865	0.642	0.611	0.712	0.666	0.579	**0.563**
80.3-90.12	0.756	**0.493**	0.528	0.695	0.606	0.529	0.600
90.3-00.12	**1.149**	1.504	1.493	1.255	1.276	1.236	1.252
70.3-90.12	0.849	0.623	0.621	0.722	0.691	**0.596**	0.617
70.3-00.12	0.942	0.872	0.865	0.848	0.852	**0.796**	0.820
80.3-00.12	0.965	0.998	1.016	0.962	0.956	**0.924**	0.948
70.3-03.12	0.945	0.873	0.875	0.850	0.833	**0.801**	0.802

Retail Sales

Period	VAR(4)	VARF					
		VAR(4) + PC	VAR(BIC) + PC	VAR(4) + PLS2	VAR(BIC) + PLS2	VAR(4) + SPLS2	VAR(BIC) + SPLS2
70.3-80.12	0.748	0.543	0.499	0.586	0.539	0.523	**0.484**
80.3-90.12	0.830	0.721	0.728	0.684	0.611	**0.603**	0.610
90.3-00.12	1.068	1.053	1.049	0.944	0.964	**0.930**	0.948
70.3-90.12	0.828	0.675	0.658	0.658	0.616	0.580	**0.556**
70.3-00.12	0.887	0.759	0.736	0.727	0.713	0.711	**0.669**
80.3-00.12	0.927	0.842	0.842	0.761	0.743	**0.726**	0.746
70.3-03.12	0.888	0.761	0.739	0.728	0.721	0.713	**0.712**

Average RMSE

Period	VAR(4)	VARF					
		VAR(4) + PC	VAR(BIC) + PC	VAR(4) + PLS2	VAR(BIC) + PLS2	VAR(4) + SPLS2	VAR(BIC) + SPLS2
70.3-80.12	0.815	0.598	0.566	0.564	0.528	0.460	**0.441**
80.3-90.12	0.799	0.574	0.608	0.578	0.502	**0.492**	0.513
90.3-00.12	1.130	1.254	1.250	1.043	1.063	**1.024**	1.038
70.3-90.12	0.835	0.621	0.619	0.584	0.550	0.490	**0.489**
70.3-00.12	0.896	0.749	0.742	0.661	0.651	0.618	**0.615**
80.3-00.12	0.930	0.834	0.856	0.760	0.724	**0.716**	0.730
70.3-03.12	0.906	0.756	0.751	0.672	0.651	0.636	**0.628**

Source: Authors' calculations. The table shows the ratio of RMSE of VAR(4), VAR(4) and VAR(BIC) +PC, +PLS2 and +SPLS2 over the benchmark model for h=24. Bold figures indicate the best forecasting method for each subsample. The bottom panel shows the average RMSE.

Table 4.6
AR(BIC)+PLS2 factors. RMSE, h=1

Industrial Production

Period	PLS2		SPLS2	
	Option a (k=2)	Option b (k=1)	Option a (k=2)	Option b (k=1)
70.3-80.12	0.842	0.824	0.807	**0.802**
80.3-90.12	0.713	0.757	0.696	0.751
90.3-00.12	0.917	0.889	0.908	0.887
70.3-90.12	0.840	0.845	0.825	0.830
70.3-00.12	0.860	0.863	0.845	0.850
80.3-00.12	0.791	0.817	0.788	0.812
70.3-03.12	0.862	0.866	0.847	0.856

Total Employment

Period	PLS		SPLS	
	Option a (k=2)	Option b (k=1)	Option a (k=2)	Option b (k=1)
70.3-80.12	0.826	0.869	0.807	0.860
80.3-90.12	0.760	0.834	**0.758**	0.833
90.3-00.12	1.010	1.006	1.027	1.015
70.3-90.12	0.824	0.872	0.812	0.874
70.3-00.12	0.842	0.886	0.834	0.887
80.3-00.12	**0.810**	0.869	0.813	0.869
70.3-03.12	0.877	0.910	0.868	0.916

Personal Income

Period	PLS		SPLS	
	Option a (k=2)	Option b (k=1)	Option a (k=2)	Option b (k=1)
70.3-80.12	0.852	0.831	0.837	**0.813**
80.3-90.12	0.903	0.942	**0.870**	0.942
90.3-00.12	0.928	0.938	0.927	0.937
70.3-90.12	0.877	0.889	**0.865**	0.876
70.3-00.12	0.907	0.920	**0.901**	0.913
80.3-00.12	0.926	0.946	**0.924**	0.946
70.3-03.12	0.909	0.926	**0.904**	0.920

Retail Sales

Period	PLS		SPLS	
	Option a (k=2)	Option b (k=1)	Option a (k=2)	Option b (k=1)
70.3-80.12	0.899	0.980	**0.896**	0.934
80.3-90.12	0.930	0.923	0.930	0.922
90.3-00.12	1.047	1.043	**1.040**	1.042
70.3-90.12	0.917	0.953	**0.917**	0.934
70.3-00.12	**0.940**	0.976	0.942	0.955
80.3-00.12	0.964	0.965	0.963	**0.962**
70.3-03.12	0.968	0.982	0.959	0.948

Source: Authors' calculations. The table shows the ratio of RMSE of AR(BIC)+PLS2 and AR(BIC)+SPLS2 over the benchmark model for h=1. Bold figures indicate the best forecasting method for each subsample.

Table 4.7
AR(BIC)+PLS2 factors. RMSE, h=6

Industrial Production

Period	PLS		SPLS	
	Option a (k=2)	Option c (k=1)	Option a (k=2)	Option c (k=1)
70.3-80.12	0.440	0.462	0.407	0.431
80.3-90.12	0.762	0.762	0.765	0.759
90.3-00.12	0.689	0.673	0.679	0.682
70.3-90.12	0.561	0.577	0.542	0.562
70.3-00.12	0.582	0.586	0.560	0.573
80.3-00.12	0.766	0.743	0.771	0.743
70.3-03.12	0.620	0.646	0.611	0.631

Total Employment

Period	PLS		SPLS	
	Option a (k=2)	Option c (k=1)	Option a (k=2)	Option c (k=1)
70.3-80.12	0.580	0.538	0.586	0.496
80.3-90.12	0.609	0.630	**0.607**	0.621
90.3-00.12	0.990	0.925	0.979	0.962
70.3-90.12	0.584	0.584	0.596	0.577
70.3-00.12	0.624	0.613	0.630	0.606
80.3-00.12	0.677	0.689	**0.674**	0.682
70.3-03.12	0.704	0.682	0.707	0.671

Personal Income

Period	PLS		SPLS	
	Option a (k=2)	Option b (k=1)	Option a (k=2)	Option b (k=1)
70.3-80.12	0.582	0.562	0.567	**0.512**
80.3-90.12	0.756	0.842	**0.752**	0.869
90.3-00.12	0.900	0.986	0.894	0.983
70.3-90.12	0.706	0.716	**0.674**	0.705
70.3-00.12	0.758	0.787	**0.739**	0.786
80.3-00.12	0.824	0.945	**0.820**	0.947
70.3-03.12	0.783	0.825	0.765	0.822

Retail Sales

Period	PLS		SPLS	
	Option a (k=2)	Option c (k=1)	Option a (k=2)	Option c (k=1)
70.3-80.12	0.482	0.519	0.443	0.457
80.3-90.12	0.624	**0.606**	0.613	0.613
90.3-00.12	0.949	0.966	**0.947**	0.989
70.3-90.12	0.565	0.579	0.541	0.545
70.3-00.12	0.648	0.652	**0.632**	0.644
80.3-00.12	0.751	**0.725**	0.744	0.736
70.3-03.12	0.656	0.663	0.640	0.656

Source: Authors' calculations. The table shows the ratio of RMSE of AR(BIC)+PLS2 and AR(BIC)+SPLS2 over the benchmark model for h=6. Bold figures indicate the best forecasting method for each subsample.

Table 4.8
AR(BIC)+PLS2 factors. RMSE, h=12

Industrial Production

Period	PLS		SPLS	
	Option a (k=1)	Option b (k=1)	Option a (k=1)	Option b (k=1)
70.3-80.12	0.232	0.246	0.197	**0.194**
80.3-90.12	0.652	0.547	0.637	0.578
90.3-00.12	0.813*	0.987*	0.887	0.988*
70.3-90.12	0.381	**0.360**	0.375	0.374
70.3-00.12	0.419	**0.415**	0.420	0.430
80.3-00.12	0.689	**0.642**	0.693	0.670
70.3-03.12	0.484	0.474	0.484	0.487

*k=2

Total Employment

Period	PLS		SPLS	
	Option a (k=1)	Option b (k=1)	Option a (k=1)	Option b (k=1)
70.3-80.12	0.406	0.387*	0.376	0.369*
80.3-90.12	**0.444**	0.542	0.465	0.527
90.3-00.12	1.045	1.014	1.058*	1.032
70.3-90.12	0.434	0.463	0.446	0.459
70.3-00.12	0.514	0.532	0.536	0.530
80.3-00.12	0.613	0.672	**0.653**	0.664
70.3-03.12	0.606	0.585	0.621	0.581

* k=2

Personal Income

Period	PLS		SPLS	
	Option a (k=1)	Option b (k=1)	Option a (k=1)	Option b (k=2)
70.3-80.12	0.530*	0.520*	0.454	**0.426**
80.3-90.12	0.695	0.838	0.702	0.816
90.3-00.12	0.974	1.064*	1.008	1.030
70.3-90.12	0.615	0.669	0.604	0.592
70.3-00.12	0.709	0.786*	0.713	0.728
80.3-00.12	0.829	0.954*	0.851	0.950
70.3-03.12	0.740	0.807*	0.743	0.749

Retail Sales

Period	PLS		SPLS	
	Option a (k=1)	Option b (k=1)	Option a (k=1)	Option b (k=1)
70.3-80.12	0.486	0.529*	0.431	0.449*
80.3-90.12	0.587	0.613	0.558	0.546
90.3-00.12	**0.913**	0.970	0.929*	0.961
70.3-90.12	0.540	0.580	0.498	0.525
70.3-00.12	0.646	0.664	0.613	0.632
80.3-00.12	0.761	0.754	0.734	0.732
70.3-03.12	0.650	0.673	**0.617**	0.641

Source: Authors' calculations. The table shows the ratio of RMSE of AR(BIC)+PLS2 and AR(BIC)+SPLS2 over the benchmark model for h=12. Bold figures indicate the best forecasting method for each subsample.

Table 4.9
AR(BIC)+PLS2 factors. RMSE, h=24

Industrial Production

Period	PLS		SPLS	
	Option a (k=1)	Option b (k=1)	Option a (k=1)	Option b (k=1)
70.3-80.12	0.337	**0.246**	0.271*	0.292*
80.3-90.12	0.401	0.547	0.405	0.464
90.3-00.12	0.946	0.987*	0.774	0.770
70.3-90.12	0.370	0.360	0.340*	0.358*
70.3-00.12	0.438	0.415	0.421*	0.430
80.3-00.12	0.547	0.642	**0.527**	0.561
70.3-03.12	0.516	0.474	0.503	0.456

*k=2

Total Employment

Period	PLS		SPLS	
	Option a (k=1)	Option b (k=1)	Option a (k=1)	Option b (k=1)
70.3-80.12	0.443	0.387	0.381	**0.375**
80.3-90.12	0.398*	0.542*	0.378	0.377
90.3-00.12	0.885*	1.014*	0.885*	0.850*
70.3-90.12	0.436	0.463	0.382	**0.377**
70.3-00.12	0.536	0.532*	0.490	0.490
80.3-00.12	0.581	0.672*	0.586	**0.563***
70.3-03.12	0.653	0.585	0.622	**0.521**

*k=2

Personal Income

Period	PLS		SPLS	
	Option a (k=2)	Option b (k=2)	Option a (k=2)	Option b (k=2)
70.3-80.12	0.651*	0.520*	0.559	0.564
80.3-90.12	0.716*	0.838	0.619	0.621
90.3-00.12	1.180	1.064	1.119*	1.117
70.3-90.12	0.695*	0.669	0.611	0.604
70.3-00.12	0.843	0.786	0.774*	0.786*
80.3-00.12	0.966	0.954	0.919*	0.948
70.3-03.12	0.884*	0.807	0.823*	0.790*

*k=1

Retail Sales

Period	PLS		SPLS	
	Option a (k=2)	Option c (k=2)	Option a (k=2)	Option c (k=2)
70.3-80.12	0.565	0.529	0.510	0.553
80.3-90.12	0.634*	0.613	0.606	0.605*
90.3-00.12	0.877*	0.970*	**0.824***	0.826*
70.3-90.12	0.651	0.580	0.572	0.597
70.3-00.12	0.720*	0.664	0.676	0.691
80.3-00.12	0.722*	0.754*	0.703*	0.713*
70.3-03.12	0.725*	0.673	0.674*	0.691

*k=1

Source: Authors' calculations. The table shows the ratio of RMSE of AR(BIC)+PLS2 and AR(BIC)+SPLS2 over the benchmark model for h=24. Bold figures indicate the best forecasting method for each subsample.

Chapter 5

Selecting and Combining Experts for Survey Forecasts

5.1 Introduction

The surveys of expert forecasts have proven to be relevant in the formation of macroeconomic expectations of economic agents. Although the forecast of the individual experts are made available by the corresponding institutions, the public usually focuses on a summary measure as the mean or median values of the forecasted variables, which generally, are also provided. The empirical finding that simple forecast combinations, in particular the equal weights average, tend to outperform individual forecasts or more sophisticated schemes is one of the reasons that may explain this practice (Clemen, 1989; Makridakis et al., 1982; Stock and Watson, 2001 and Genre et al., 2013; among others).

The idea of examining combination methods that could better exploit the information provided by multiple forecasts of the same variable has motivated a large literature on forecast combination (Bates and Granger, 1969; Marcellino, 2004; Timmermann, 2006; among others). The evidence has demonstrated that forecast combinations improve the performance of individual forecasts. Hence, recent research has focused on the problem of the optimal combination of forecasts. There are several methods proposed in the literature to estimate combination weights. From the seminal paper by Bates and Granger (1969), to more recent sophisticated alternatives as, for instance, factor methods (Poncela et al., 2011) and Bayesian shrinkage combinations (Diebold and Pauly, 1990) there is a large volume of literature that has focused on this issue.

In general, to forecast the target variable y_{t+h} the previous methods combine the available set of predictors or individual forecasts $y^1_{t+h|t},\ldots,y^N_{t+h|t}$, assigning a positive weight to all of them. In this chapter we are interested in performing a selection of informative experts from a survey, discarding some individual forecasts. Some forms of trimming have been proposed (Stock and Watson, 2004; Granger and Jeon, 2004 and Aiolfi and Favero, 2005). Its simplest form is to discard α% of the lowest and highest values of the forecasts and then take the average of the remaining ones. When the number of forecasts to combine is high, the trimming technique can be quite aggressive (see, for instance, Samuels and Sekkel 2013, in the context of combination of forecast from models, instead of surveys). Our proposal is somewhat different, because it tries to remove the forecasts with redundant information or low predictive power about the target; not extreme ones. Conflitti et al. (2012) in their approach for point forecast of the European Central Bank Survey of Professional Forecasters (ECB SPF) is the closest to our scheme. Using the well-known Stock and Watson database, Fuentes et al. (2014) find that the selection of predictors achieved by the sparse partial least squares (SPLS) improves the forecast efficiency compared to the widely used competing models, including the least angle regression pure selection procedure.

Poncela et al. (2011) analyzed several multivariate techniques to combine the expert forecasts. For each of the six United States (U.S.) Business Indicators considered from the Survey of Professional Forecasters (SPF) for the period 1991-2008, they find that taking into account the target variable y_{t+h} in the process of dimension reduction, the combination outperforms the standard benchmark (simple average of individual forecasts) and, in particular, partial least squares (PLS) provides a good forecasting performance.

We explore the empirical performance of different ways to combine forecasts and to combine selected forecasts from surveys. We use the individual forecasts for U.S. economic variables, collected by the Philadelphia Federal Reserve Bank's SPF for the period 1991-2012. In addition, we divide the entire sample into three periods to evaluate the robustness and the business cycle sensitivity of the results.

In particular, we propose to investigate the usefulness of SPLS, a technique that allows selecting and combining the informative predictors for a forecasted target.

The chapter is organized as follows. Section 5.2 briefly presents the different combination and selection methods focused on the empirical analysis. Section 5.3 describes some relevant features of the SPF dataset. Section 5.4 presents the empirical applications and the forecasting results. Finally, Section 5.5 concludes.

5.2 Forecast Combination and/or Selection Methods

In this section, we briefly describe the different techniques to combine the forecasts based on survey data that we use in the empirical application. Some approaches include the information from the full panel of forecasters in the combination: ordinary least squares (OLS) and PLS; while others perform a selection of forecasters before combining them: trimmed mean, LASSO and SPLS.

For the implementation of LASSO by LARS, we will retain the first k selected forecasts for their inclusion in an OLS regression. We also apply the SPLS algorithm proposed by Chun and Keles (2010) to this dataset.

5.2.1Trimmed Mean

The trimmed mean is the mean computed by excluding a $\alpha/2\%$ of the lowest and highest values from a sample. For example, a mean trimmed by 50% has 25% of the largest forecasts and 25% of the smallest forecasts removed.

This is a measure of central tendency and it is considered a robust estimator of location for a symmetric distribution, because it reduces the influence of outliers. The median can be regarded as an extreme trimming method.

5.2.2 Partial Least Squares (PLS)

As was mentioned before, this technique constructs a scheme for extracting orthogonal latent factors based on the covariance between the predictors (X) and the dependent or forecasting variable (Y). In this particular case, $X_t = (y^1_{t+1|t},\ldots,y^N_{t+1|t})$ is an N-dimensional vector of one-step ahead forecasts of the target variable from the survey of panelists at time t, $\underset{t\times N}{X} = (\mathrm{X}_1^{'}\,\mathrm{X}_2^{'}\ldots\mathrm{X}_t^{'})'$ and $Y = (Y_1\,Y_2\ldots Y_t)'$.

5.2.3 Some Intuitions

After introducing the forecasting combination methods, we are going to motivate them through the population case of combining N experts. This will allow us to understand when and how the different methods may work. To keep the exposition as simple as possible, we will present the case of just two linear combinations of one period ahead forecast $y^i_{t+1|t}$, i = 1, …, N.

Assume that since all experts want to forecast the same target value, they can be highly correlated. The OLS combination of the experts is given by:

$$y_{t+1} = \beta_0 + \beta_1 y^1_{t+1|t} + \ldots + \beta_N y^N_{t+1|t} + \varepsilon_{t+1}. \qquad (5.1)$$

Due to multicollinearity, the uncertainty in the estimation of the β's can be quite high. Therefore, using them for forecasting can produce unstable forecasts, especially in the turning points. For instance, assume that all experts have been quite collinear in the past but a few of them foresee a recession while the remaining ones do not.

The weights and β's for each of the methods mentioned above are different. For the mean or average of forecasters $\beta_i = 1/N$, i=1, ... , N and $\beta_0 = 0$, while for the trimmed mean $\beta_i = 1/[(1-\alpha)N]$, if forecaster i is not trimmed and 0 otherwise, where 1-α is the percentage of central observations considered for the estimation of the mean, [(1-α)N] stands for the positive integer closest to (1-α)N and as in the previous case $\beta_0 = 0$.

Notice that with the average and the trimmed average we go from N individual forecasters to just 1 final combined forecast through:

$$f_t = \sum_{i=1}^{N} \beta_i y^i_{t+1|t}. \qquad (5.2)$$

In spite of its simplicity, the average of forecasters has been widely used and has been continuously analyzed (see, for instance, Clements and Harvey, 2009; Timmermann, 2006 and Genre et al., 2013; for some recent papers). As it is well known, the average of forecasts works well when the variance associated to each forecaster is the same (homogeneous forecasts), irrespective of the correlation among them.

In the case of heterogeneous forecasts, because some of the forecasts may be biased or even if we consider only unbiased forecast, the variance of each one of them may be different, the selection by the trimmed mean might work well because it can discard biased forecasts and inefficient forecasters in the sense of high variance. A particular case that may be of interest for macroeconomics is the case of a bimodal distribution of the cross section of forecasters. For instance, this might correspond to the previous mentioned case of entering into a recession that only a few forecasters may anticipate. In this case, the trimmed mean may not work as desired if we discard the forecasters that foresee the recession.

However, in cases where the forecast error from expert i is quite large, the trimmed mean might be a good solution. When the number of experts is large, it can be a good method for choosing the central experts of the distribution. In our forecasting problem, the experts were already selected by their continuity in participating in the survey and therefore our sample is already trimmed in some sense.

Equation (5.1) can be approximated by selecting some of the experts as LASSO does. Another alternative is to rely on a small set of linear combinations of the experts or factors that capture common behavior of the experts.

Concerning the first approach, selecting some of the experts might yield better forecasting results if there is a group of "better" experts, that is, experts that systematically have a small MSE. If this is the case, it might be worth it selecting them and discarding the remaining ones.

As regards exploiting the common information shared by the experts, equation (5.1), seen as the combination rule for the experts, might be seen as the result of a two step procedure where first the experts are combined as follows

$$f_t^j = \sum_{i=1}^{N} \omega_{ij} y_{t+1|t}^i, \quad \text{j=1, …, k.} \tag{5.3}$$

Then, the k linear combinations of the forecasts are regressed over the target

$$y_{t+1} = \gamma_0 + \sum_{j=1}^{k} \gamma_j f_t^j + \varepsilon_{t+1} \tag{5.4}$$

Note that (5.4) should be estimated with information up to time t to generate true ex ante forecasts. Observe that (5.3) and (5.4) are equivalent to (5.1).

In cases where the experts are highly collinear, they may belong to a subspace of lower dimension than the number of experts. In this situation, finding as many components as the dimension of the subspace spanned by the experts might result in a lower RMSE. If these k linear combinations are oriented towards the forecasting target, the factors are obtained by

PLS were in the simplest case of one common component $\omega_i \propto Cov(y^i_{t|t-1}, y_t)$ and $\gamma_1 = \frac{Cov[y_t, f^{pls}_{t-1}]}{V[f^{pls}_{t-1}]}$.

The flexibility of PLS and SPLS is that we can increase the number of components and cover cases of heterogeneous forecasters. As regards the case of PLS recall that the weights take into account the covariance with the target, giving more weight to those forecasters more correlated with the target. The first PLS component takes into account the univariante effect on the target of each forecaster. We seek for directions of high variance as well as high correlation with the target.

To give some intuitions where the PLS approach might work well, consider the case where the majority of forecasters are highly collinear but the small group that foresee a recession, although the majority do not. In this case, the first PLS component gives weights to all forecasters as usual. The second PLS component, orthogonal to the first one, can capture the variation in the small group of forecasters that disagree if this variation is on the direction of the target. Notice that this effect cannot be captured by the average forecast.

However, there are cases were the weights ω_i are not too high and their estimation uncertainty might not compensate their contribution to the final forecast. In this instance, we can obtain linear combinations of the experts with a smaller MSE by setting some of the weights to zero.

Additionally, there are other situations where sparsity might yield better results. In PLS, we give some weight to all the experts. If there is a small group of them that foresee the recession, their weight is diluted by the number of experts, perhaps not giving them enough weight. If we could introduce some sparsity by zeroing out some redundant experts or not very informative ones, the group that anticipates the turning point might be not too diluted and capture the change in the economy. We might need more than 1 sparse component to pick up the previous behavior.

5.3 Empirical Application

To explore how selecting and combining expert forecasts works in practice, we employ the dataset generated by the Survey of Professional Forecasters (SPF). The SPF is conducted by the Federal Reserve Bank of Philadelphia since 1990, when they took over the American

Statistical Association (ASA) and the National Bureau of Economic Research (NBER) project started in 1968.

The survey is conducted on a quarterly basis and asks the forecasters to provide projections for the next five quarters: the quarter in which the survey is conducted and subsequent four quarters. Advanced reports of government statistical agencies such as the U.S. Bureau of Economic Analysis (BEA) and the U.S. Bureau of Labor Statistics (BLS) of the previous quarter are reported in the survey questionnaire sent to the panelists.

We use the one-step-ahead forecast of the consumer price index (CPI) and five variables of the U.S. Business Indicators group: nominal gross domestic product (NGDP), 3 month Treasury bill rate (TBILL), AAA corporate bond yield (BOND), civilian unemployment rate (UNEMP) and housing starts (HOUSING). We assume that the logarithm of NGDP is integrated of order 1 and that the levels of the remaining series are difference stationary.

For all variables, the panel of forecasters used spanned the period from the first quarter of 1991 to the fourth quarter of 2012. Because the dataset is particularly unbalanced, due to the entry and exit of forecasters, we apply two types of pre-treatments on the information from the SPF. First, as in Poncela et al. (2011) we reduce the pool of panelists by selecting those forecasters who met the following criteria: (i) they have been on the panel at least seven years and (ii) they have no more than four consecutive quarters without an active participation in the survey. Second, missing observations are filled with the marginal mean of the forecast up to the previous period given by the respondent. If the missing observations occur in the first period of the estimation sample, they are filled with the non-revised or first estimated data of the previous period.

To perform a sensibility analysis and a robustness check, we divide the full sample in three periods. One forecasting sample includes only a period of economic expansion. The remaining two samples contain recession periods according to the NBER dating. Table 5.1 summarizes the estimation and forecasting samples.

The period of projection of the first sample spanning from the first quarter of 2000 to the fourth quarter 2003 contains the dot com recession. In the third sample, the forecast sample covers from the first quarter 2007 to the fourth quarter 2012 including the last deep recession; while for the second sample is an expansionary period covering from the first quarter in 2005 to the fourth quarter in 2007.

The entry and exit of panelists makes it extremely difficult to have a long sample for estimating the competing methods. In this case, it is not possible to perform an analysis for the full sample. The number of panelists considered oscillates between 14 and 19, depending on the variable and the subsample (see Table 5.2).

5.4 Forecast Results

To compare the forecast accuracy of the trimmed mean, OLS, PLS, LARS and SPLS combination schemes relative to the equal weighted benchmark, we use the relative root mean squared forecast error (RMSE):

$$\mathrm{Re}lative\ RMSE = \frac{RMSE\ (method)}{RMSE\ (average)}. \tag{5.5}$$

An entry of less than one implies that the combination scheme outperforms the simple average forecast.

For the trimmed mean, we employ two values of α: 20% and 50%. As regards LARS, we will retain the first 5 and 10 selected forecasts for their inclusion in an OLS regression. The SPLS approach is implemented considering different values for the sparsity parameter (λ) = 0.2, 0.4, 0.6 and 0.8. The number of components considered for PLS and SPLS are k=1, 2 and 3.

5.4.1 Relative Performance of the Competing Methods

The forecasting results for the six variables and the three forecasting subsamples considered are shown in Tables 5.3 and 5.4. The comparisons performed suggest several notable features. First, they show that in all cases there is a method that works better than the simple average. We find predictive gains in techniques that perform a forecaster's selection: trimmed mean, LARS and SPLS. Second, SPLS yields the best forecasting performance in 72.2% of cases, and its accuracy is similar to the best alternative model in 16.7% of the remaining ones. Third, SPLS provides the best forecasting results for the variables BOND, CPI and UNEMP in all subsamples considered. Fourth, we find forecast improvements of SPLS with respect to the benchmark for TBILL. For the first subsample (00.1-03.4) SPLS provides the best performance and for the other two subsamples, in which LARS (10) gives the best results, it performs quite well and its accuracy is close to that of LARS. Fifth, there are two particular cases for which the benchmark has proven to be hard to beat: (i) the subsample (05.2001-07.2006) for NGDP and (ii) the subsample (07.2001-12.2004) for

HOUSING, in both cases the trimmed mean (α=50%) is the only method able to produce accuracy gains over the simple average.

Note that in cases where the average of forecasters works well in comparison with OLS, 1 component (PLS or SPLS) should also work well. In this case selecting more components is not a wise solution since we will be approaching OLS. This can be seen, for instance, in the table for unemployment. The reason why PLS or SPLS might work better than the average of forecasts is that we can give more weight to those forecasters more correlated with the target. The key issue is not with extreme values (trimmed mean) but giving more weight to the most correlated forecasters with the target. On the contrary, if OLS outperforms the average of forecasts, then choosing more PLS or SPLS components should be advisable.

In general, the best performing SPLS models have just one component and a high degree of sparsity (λ=0.8). However, the number of panelists selected depends on the variable and on the period measured. For HOUSING and NGDP the number of informative forecasters is reduced significantly in all periods analyzed. For the rest of variables, the panel composition changes with the subsample.

5.4.2 Sensitivity Results to Business Cycle

5.4.2.1 RMSE

As can be seen from Tables 5.3 and 5.4, the relative performance of the competing models over all the subsamples and variables is quite similar. The inclusion of the financial crisis period of 2008-2009 in the third subsample does not seem to produce a notable effect in their forecasting performance with respect to the average of forecasters, except for HOUSING[1]. In this later case, the influence of the housing sector crash of 2006-2007 –which is considered the worst in the U.S history and the root of the global crisis- and its sluggish recovery, could explain the observed decline in the forecast accuracy of the different methods.

It is important to highlight that the size of the estimation sample varies among subsamples, as well as the forecasters involved in them; thus, the uncertainty of estimation also differs among them.

[1] In the case of BOND, it is observed deterioration in the accuracy of the forecast in relation to the second sample.

To evaluate in more detail the forecasting results during the pre-crisis and crisis periods, the second subsample was extended until the fourth quarter of 2009. For this subsample, the first estimation period starts in 1996:01 and ends at 2004:04 and the forecasting period covers from 2005:01 until 2009:04. Table 5.5 reports the relative RMSE results.

The results for the extended sample 2 evidence that a number of the considered schemes outperform the equal weighted combination. In particular, with the exception of NGDP, the best performing method across variables is SPLS. The SPLS method improves over the benchmark and it yields gains between 16.7% and 37.1%; the greatest reduction in the RMSE was achieved by CPI. The good performance of some of the combination schemes for inflation has been found for the ECB SPF, over normal business cycle conditions (Genre et al., 2013). In the case of the Harmonised Index of Consumer Prices (HICP) for the Euro area, the authors link the positive results of the combination strategies to the correction of a persistent downward bias in the inflation forecasts.

Table 5.6 compares the RMSE of the extended subsample splitting the prediction errors in two groups; those related to the "Non-Crisis" period and a second group corresponding to the time period 2008:01-2009:02, defined as the recession by the NBER Business Cycle Dating Committee. For CPI, NGDP, TBILL and UNEMP, we find a relative better performance with respect to the average forecasts in the crisis period than in the non-crisis period. BOND and HOUSING present the opposite behavior.

A stylized fact in the literature is that housing investment leads the business cycle. As was mentioned before, the beginning of the crisis period of the housing sector is prior to that used in this comparison, even though it still represents a deceleration period in which the results show a high sensitivity.

During the recession period, the squared forecast errors of the SPLS model reported an increase across variables, reflecting the increased uncertainty associated with the financial crisis. For TBILL this phenomenon was observed in advance from the second quarter of 2007 to the third quarter of 2008. As previously noted, in the case of HOUSING, forecast errors also observed an increase during the year 2006. For NGDP, BOND and CPI the forecast errors behavior also seems to have changed in 2006 (see, Figure 5.1).

5.4.2.2 The Degree of Sparsity

The deteriorating macroeconomic environment resulting from the crisis is evidenced in the degree of sparsity observed along the prediction period. For NGDP and UNEMP, the number of forecasters included in the combinations began to increase in the second and third quarter of 2008, respectively, and their positive trend was maintained until the end of the subsample (see, Figure 5.2).

For TBILL and HOUSING, the number of forecasters began to grow since the second quarter of 2005 and peaked in June 2006. For the first variable, the number of forecasters remains in its peak value until the end of the subsample, while for HOUSING after showing a declining trend until the last quarter of 2007, it reports an increase during 2008. Recall that the behavior of TBILL is related to the monetary policy and HOUSING reflects the state of the business cycle earlier than other indicators (see, for instance, Leamer, 2009). Finally, the number of forecasters for CPI and BOND remains unchanged during the crisis period showing a slight increase at the end of the sample.

5.5 Conclusions

As was stated by Clemen (1989) and confirmed by a large number of empirical studies “Forecast accuracy can be substantially improved through the combination of multiple individual forecasts”. However, the question of how to better combine the available forecasts is still an open issue.

In this chapter, we compare the performance of different techniques to combine all individual forecasts of the panel of forecasters and to combine only selected single forecasts from the SPF. The empirical results suggest that combination schemes that performs a forecaster’s selection yield predictive gains over the widely used summarizing measure, the simple average of the survey participants. In particular, the selection process implemented by the SPLS method provides a good prediction performance. The final SPLS forecast combinations are sparse, which implies that forecasters with redundant information or small predictive power over the target are removed from the panel, reducing the estimation error and, therefore, improving the forecast efficiency.

Considering the four subsamples analyzed (including the extended sample 2) for the period 1991-2012, we find that the SPLS model outperforms the alternative methods in almost 75% of the cases. It is important to highlight that the extended sample 2 includes specifically the crisis period 2008-2009. According to the robustness of the results obtained in the different subsamples and the detailed results for the fourth subsample, the model was able to capture the changes in the behavior of the variables resulting from the financial crisis and performed well during its development, with the exception of HOUSING and BOND. We believe that the prolonged stagnation of the housing sector, the severity of the crisis and the sluggish economic recovery have influenced the deterioration of the performance of the forecasting models of these latter variables with respect to the average of forecasters.

The economic uncertainty caused by the financial crisis can be evidenced in the change that occurs in the number of panelists who are selected by the model over the prediction period 2005-2009. For NGDP and UNEMP the number of the panelists selected to build the final combination increased during the recession period 2008-2009. For HOUSING, a similar behavior was observed temporarily in accordance with the cycle on the sector and was repeated, with less intensity, during the recession of the economy.

References

Aiolfi, M., Capistrán, C. and A. Timmermann (2011). Forecast Combinations. In *Oxford Handbook of Economic Forecasting,* Clements M. and Hendry, D. (eds). Oxford University Press; 355-388.

Aiolfi, M. and C. Favero (2005). " Model Uncertainty, Thick Modelling and the Predictability of Stock Returns". *Journal of Forecasting*. Vol 24(4): 233–254.

Bates, J. and C. Granger (1969). "The Combination of Forecasts". *Operational Research Quarterly*, 20 (4): 451-468.

Clemen, R. (1989). "Combining Forecasts: A Review and Annotated Bibliography". *International Journal of Forecasting*, 5 (4): 559-583.

Clements, M and D. Harvey (2009). Forecast Combination and Encompassing. In *Palgrave Handbook of Econometrics, Applied Econometrics*, Mills T. and Patterson, K. (eds). Oxford Palgrave, (2):169-198.

Chun H. and S. Keles (2010). "Sparse Partial Squares Regression for Simultaneous Dimension Reduction and Variable Selection", *Journal of the Royal Statistical Society, Series B, Statistical Methodology*, 72(1): 3-25.

Conflitti, C., De Mol, C. and D. Giannone (2012). "Optimal Combination of Survey Forecasts". ECARES Working Paper 2012-023.

Diebold, F. and P. Pauly (1990). "Structural Change and the Combination of Forecasts". *International Journal of Forecasting,* 6 (4): 503-508.

Fuentes, J., Poncela, P. and J. Rodríguez (2014). "Sparse Partial Least Squares in Time Series for Macroeconomic Forecasting". *Journal of Applied Econometrics,* Forthcoming.

Genre, V., Kenny, G., Meyler, A. and A. Timmermann (2013). "Combining Expert Forecast: Can Anything Beat the Simple Average?". *International Journal of Forecasting*, 29:108-121.

Granger, C. and Y. Jeon (2004): "Thick modeling". *Economic Modelling*, 21: 323-343.

Groen, J. and G. Kapetanios (2008). "Revisiting Useful Approaches to Data-Rich Macroeconomic Forecasting", Federal Reserve Bank of New York Staff Report 327.

Hastie T., Tibshirani, R. and J. Friedman (2008). *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer Series in Statistics, 2nd. edition.

Kelly, B and S. Pruitt (2012). "The Three-Pass Regression Filter: A New Approach to Forecasting Using Many Predictors", Chicago Booth Paper No. 11-19. Fama-Miller Center for Research in Finance.

Leamer, E. (2009). "*Macroeconomic Patterns and Stories. A Guide for MBAs*" Springer-Verlag Berlin.

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E. and R. Winkler (1982). "The Accuracy of Extrapolation (time series) Methods: Results of a Forecasting Competition". *Journal of Forecasting*, 1: 111-153.

Marcellino M. (2004). "Forecast Pooling for European Macroeconomic Variables". *Oxford Bulletin of Economics and Statistics*, 66(1): 91-112.

Mincer, J. and Zarnowitz, V. (1969). The Evaluation of Economic Forecasts. In *Economic Forecasts and Expectations:* Analysis of Forecasting Behavior and Performance. Mincer, J. (ed). New York: National Bureau of Economic Research: 1-46.

Poncela, P., Rodríguez, J., Sánchez-Mangas R. and E. Senra (2011). "Forecast Combination through Dimension Reduction Techniques". *International Journal of Forecasting,* 27: 224-237.

Samuel, J. and R. Sekkel (2013). "Forecasting with Many Models: Model Confidence Sets and Forecast Combination". Working Paper 2013-11, Bank of Canada.

Stock J. and M. Watson (2001). “A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series”. In *Cointegration, Causality, and Forecasting: Festschrift in Honour of Clive W.J. Granger,* R.F. Engle and H. White (eds). Oxford University Press: 1-44.

Stock J. and M.Watson (2004). “Combination Forecasts of Output Growth in a Seven-Country Data Set". *Journal of Forecasting*, 23: 405-430.

Tibshirani, R. (1996). “Regression Shrinkage and Selection via the Lasso”. *Journal of the Royal Statistical Society*. *Series B (Methodological)*, 58(1): 267-288.

Timmermann, A. (2006). “Forecast Combinations”. In *Handbook of Economic Forecasting*, Elliot, G., Granger, C. and A. Timmermann (eds.). North-Holland Vol I; 136-196.

Wold, H. (1966). “Estimation of Principal Components and Related Models by Iterative Least Squares”. In *Multivariate Analysis,* Krishnaiaah PR. (ed.), Academic Press: New York; 391-420.

Table 5.1
Estimation and forecasting subsamples

	Estimation subsample	Forecast subsample
M1	1991:02 to 1999:04-h	2000:01 to 2003:04
M2	1996:01 to 2004:04-h	2005:01 to 2007:04
M3	2001:01 to 2006:04-h	2007:01 to 2012:04

Table 5.2
Number of forecasters of each subsample

	BOND	CPI	HOUSING	NGDP	TBILL	UNEMP
00.1-03.4	14	14	14	16	15	16
0.5.1-07.4	16	16	17	18	16	18
07.1-12.4	14	17	15	18	16	19

Table 5.3
Relative RMSE for BOND, CPI and UNEMP. h=1

BOND. h=1

Period	Trimmed mean		OLS	PLS			SPLS			LARS	
	20%	50%		k=1	k=2	k=3	k=1	k=2	k=3	5	10
00.1-03.4	1.016	1.003	1.132	0.971	1.011	1.100	**0.846**	0.980	1.018	0.873	1.013
0.5.1-07.4	0.921	0.887	0.786	0.909	0.740	0.795	0.727	**0.713**	0.784	0.827	0.850
07.1-12.4	0.936	0.963	0.968	0.928	0.937	0.935	**0.810**	0.814	0.934	0.833	0.940

CPI. h=1

Period	Trimmed mean		OLS	PLS			SPLS			LARS	
	20%	50%		k=1	k=2	k=3	k=1	k=2	k=3	5	10
00.1-03.4	1.011	1.023	1.044	0.962	0.939	1.049	**0.899**	0.939	1.014	0.911	0.987
0.5.1-07.4	0.986	1.021	1.236	0.869	1.005	1.217	**0.843**	1.005	1.130	0.979	1.221
07.1-12.4	0.962	0.955	1.293	0.869	0.654	0.855	0.832	**0.613**	0.815	0.775	0.870

UNEMPLOYMENT. h=1

Period	Trimmed mean		OLS	PLS			SPLS			LARS	
	20%	50%		k=1	k=2	k=3	k=1	k=2	k=3	5	10
00.1-03.4	0.973	0.889	1.934	0.840	1.124	1.413	**0.839**	1.006	1.402	1.067	1.272
05.1-08.4	0.932	0.858	1.285	0.697	0.953	1.082	**0.697**	0.838	0.889	0.919	1.133
07.1-12.4	0.940	0.847	1.288	0.766	0.900	0.921	**0.766**	0.876	0.900	0.969	0.992

Source: Authors´ calculations. The table shows the ratio of RMSE of Trimmed Mean, OLS, PLS, SPLS and LARS over the benchmark model for h=1. Bold figures indicate the best forecasting method for each subsample

Table 5.4
Relative RMSE for TBILL, NGDP and HOUSING. h=1

TBILL. h=1

Period	Trimmed mean		OLS	PLS			SPLS			LARS	
	20%	50%		k=1	k=2	k=3	k=1	k=2	k=3	5	10
00.1-03.4	0.973	0.786	1.266	0.667	0.752	0.847	**0.650**	0.747	0.847	0.746	1.050
0.5.1-07.4	0.986	0.976	0.839	0.841	0.904	0.795	0.811	0.898	0.791	0.872	**0.780**
07.1-12.4	0.972	0.934	0.738	0.754	0.732	0.682	0.754	0.711	0.678	0.694	**0.655**

NGDP. h=1

Period	Trimmed mean		OLS	PLS			SPLS			LARS	
	20%	50%		k=1	k=2	k=3	k=1	k=2	k=3	5	10
00.1-03.4	1.006	1.005	2.275	**0.925**	1.128	1.318	0.926	1.135	1.301	1.406	1.565
0.5.1-07.4	1.006	**0.982**	1.785	1.306	1.370	1.751	1.235	1.365	1.719	1.601	1.655
07.1-12.4	1.005	1.015	1.548	1.040	1.010	1.077	1.035	**0.967**	1.077	0.994	1.172

HOUSING. h=1

Period	Trimmed mean		OLS	PLS			SPLS			LARS	
	20%	50%		k=1	k=2	k=3	k=1	k=2	k=3	5	10
00.1-03.4	0.996	0.980	0.824	0.841	0.777	0.823	**0.710**	0.758	0.823	0.773	0.837
0.5.1-07.4	0.952	0.967	0.796	0.906	0.742	0.807	0.774	0.700	**0.662**	0.724	0.750
07.1-12.4	0.951	**0.943**	2.136	1.549	1.460	1.681	1.407	1.463	1.657	1.581	1.743

Source: Authors´ calculations. The table shows the ratio of RMSE of Trimmed Mean, OLS, PLS, SPLS and LARS over the benchmark model for h=1. Bold figures indicate the best forecasting method for each subsample.

Table 5.5
Relative RMSE. Forecasting Sample: 2005-2009

Variable	Trimmed mean		OLS	PLS			SPLS			LARS	
	20%	50%		k=1	k=2	k=3	k=1	k=2	k=3	5	10
BOND	0.921	0.907	1.146	0.946	0.856	1.041	0.846	**0.833**	1.030	1.045	1.224
CPI	0.987	0.947	1.187	0.712	0.670	0.896	**0.629**	0.670	0.801	0.688	0.891
HOUSING	0.960	0.944	1.009	1.066	0.952	0.996	0.828	**0.804**	0.882	0.878	0.940
NGDP	1.000	**0.978**	1.697	1.055	1.147	1.492	1.043	1.063	1.425	1.241	1.705
TBILL	0.976	0.949	0.951	0.712	0.900	0.788	**0.708**	0.778	0.785	0.949	0.974
UNEMPL	0.935	0.952	1.166	0.680	0.894	0.979	**0.680**	0.877	0.974	0.992	0.952

Source: Authors´ calculations. The table shows the ratio of RMSE of Trimmed Mean, OLS, PLS, SPLS and LARS over the benchmark model for h=1. Bold figures indicate the best forecasting method for each subsample.

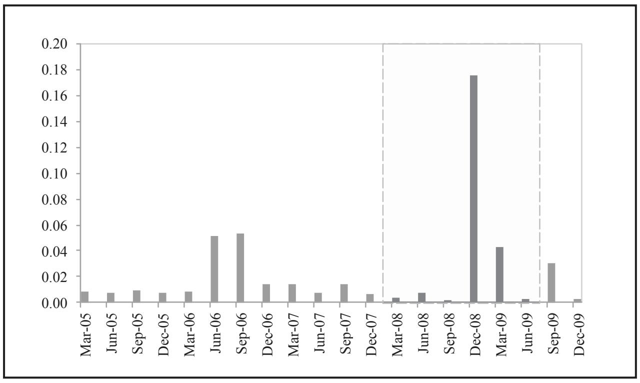
Table No. 5.6
Relative RMSE for Non-Crisis and Crisis Period

Variable	SPLS	
	Non-crisis	Crisis
BOND	0.457	1.451
CPI	0.683	0.260
HOUSING	0.563	0.873
NGDP	1.757	0.849
TBILL	0.710	0.289
UNEMP	0.777	0.414

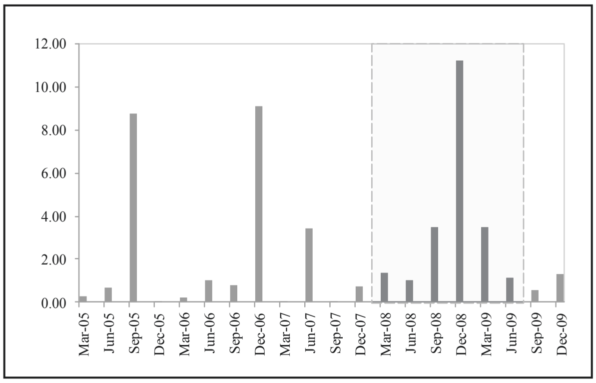
Source: Authors´ calculations. The table shows the ratio of RMSE of SPLS over the benchmark model.

Figure 5.1
Squared Forecast Errors

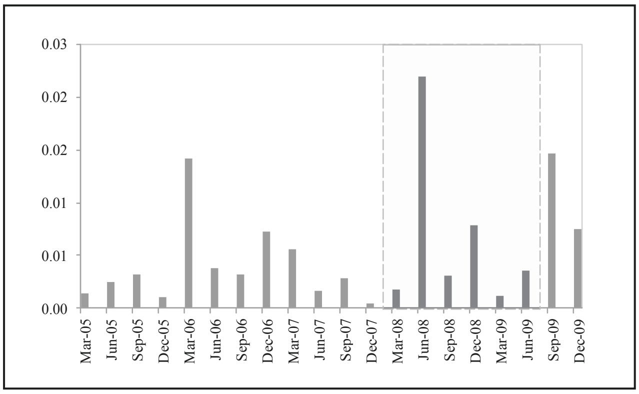
ΔBOND



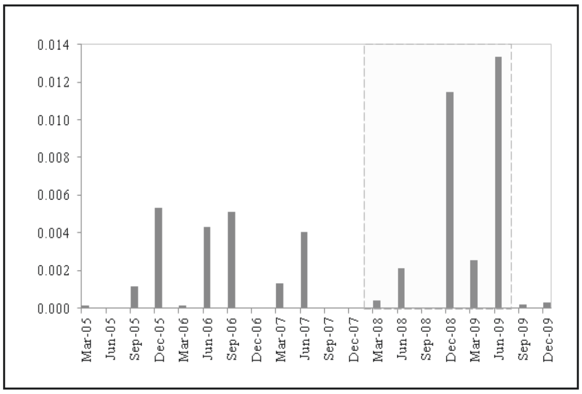
ΔCPI



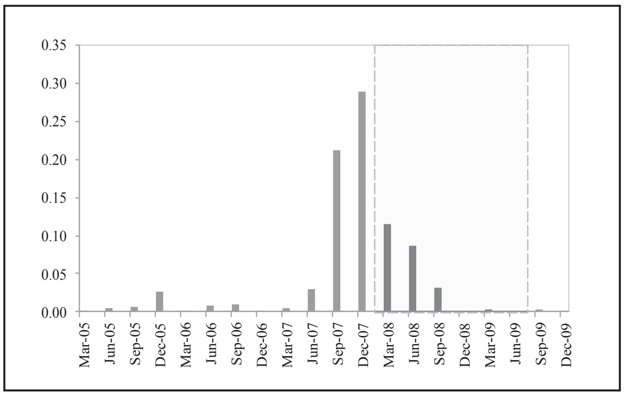
ΔHOUSING



ΔLNGDP



ΔTBILL



ΔUNEMPLOYMENT

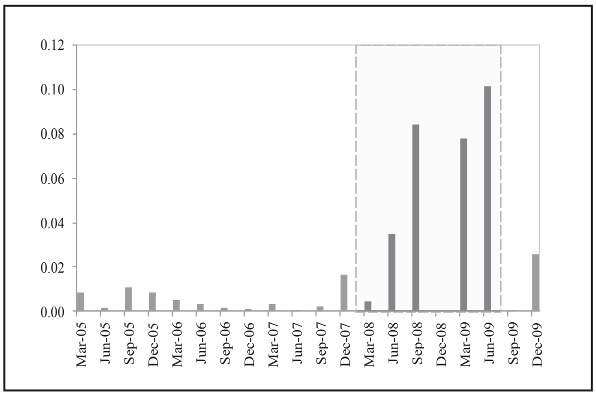
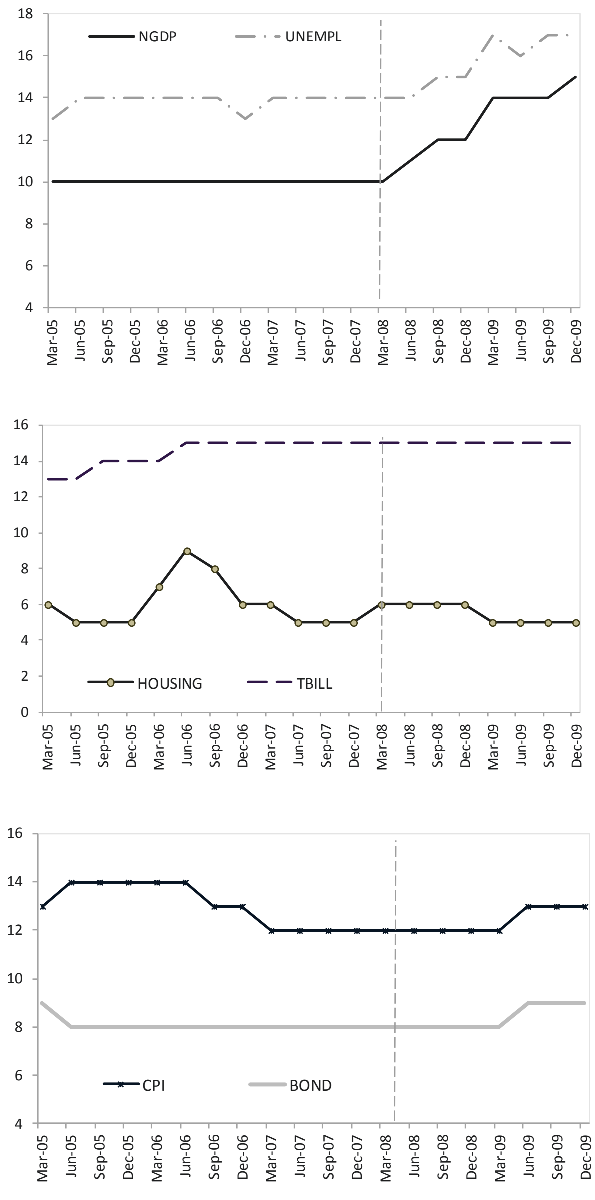


Figure 5.2
Number of Forecasters Included in the SPLS Combination



Chapter 6

Conclusions and Further Research

The empirical findings indicate that there is some room for developing the factor-based methods. In particular, the introduction of regularization to estimate the factors, based on a reduced number of relevant variables for the forecasting target, rather than using the whole dataset has been essential to improve the forecast accuracy of the discussed problems. Increasing the number of variables indefinitely does not necessarily cause better forecasting performance.

For macroeconomic forecasting with a large data set, the empirical comparisons are performed in terms of the relative Mean Square Error (RMSE) of the forecast errors, using an AR(4) process as a benchmark. We compare PLS, SPLS and the most widely used methods: principal components, targeted predictors and LARS using the well-known Stock and Watson database. The dynamic SPLS methodology improves the forecast efficiency of the competing models. Furthermore, the results of the Diebold and Mariano test suggest that the SPLS forecasts gains are statistically significant.

Among the different possibilities analyzed to apply PLS and SPLS to time series data, it seems that applying directly the PLS techniques between the target variable and the predictors yields better forecasting results. Enlarging the data set of predictors, by including the lags of the target variable in it, does not seem to be a good alternative for PLS when applied to time series data, although this is not necessarily the case when the sparse version is applied. The PLS method gives weight to all the forecasting predictors, so the dependence between the target variable and its past can be obscured if there are too many predictors. On the contrary, including the lags of the target variable explicitly on the forecasting equation seems to be the best way of capturing the dynamic behavior of the target.

The forecast performance of SPLS improves with the forecast horizon. This might reflect the fact that when the dynamics of the own lags die out, the predictive content of the cross section emerges. This is observed in most of the approaches analyzed in contrast to the pure AR(4). Taking into account that in the very short run (h=1), the forecasting results given by all the methodologies are much closer; the dynamic SPLS approaches seem to perform quite well. When the dynamic relationship is integrated through the inclusion of the lags of the target as additional predictors in the original dataset, the selection process seems to weight the relevant information for forecasting purposes appropriately. In particular, the presence of variables that have a negligible effect on the response do not lessen the participation of the lags of the target. For the updated dataset, the isolation of the AR(p) process effects, before PLS estimation, shows also a good performance at all forecasting samples.

Concerning the number of components, the best forecasting results were obtained with a small number of components: one or two.

As regards the selection of variables, the proposed method allows period by period to identify the variables that are useful for forecasting a given target. The selection performed by the SPLS model shows differences between periods of high and low uncertainty in the economic environment and between forecast horizons. According to the results, the degree of sparsity is higher when the forecast sample does not include instability periods such as the 70's. Likewise, the number of variables selected increases with the forecast horizon, because of the necessity to account for additional sources of variability to explain the target. This flexibility offers a major interpretably for the estimated factors and it can be used to distinguish periods when changes occur in the relationships among variables.

With reference to the extension the univariate implementation of PLS and SPLS methods to a multivariate framework, the results show that how potentially relevant information is added to a standard multivariate model is relevant to improve its forecast performance. Specifically, the incorporation of jointly estimation of the factors oriented toward vector of targets variables (PLS2) led to an improvement in the forecasting results of the VAR method and the VAR extended with PC factors. Furthermore, if these factors are extracted from a reduced set of relevant predictors (SPLS2) the forecasting performance is even better. The performance of the VARF augmented with PLS2 and SPLS2 factors improves with the forecast horizon.

Concerning the dynamic relationship between the response variables, it seems that the dynamic of a time series it is better capture through its own lags. Hence, the implementation of a restricted model which take into account only the own lags of each time series yields better forecasting performance.

Finally, for forecast combination from survey forecasts, the empirical comparisons are performed in terms of the relative Root Mean Square Error (RMSE) of the forecast errors, using the equal weighted combination as a benchmark. We compare SPLS with some of the most widely used methods such as trimmed mean, OLS, PLS and LARS using the Survey of Professional Forecasters (SPF). The proposed SPLS combination scheme outperforms the hard to beat average of forecasters.

The empirical results suggest that combination schemes that performs a forecaster's selection yield predictive gains over the widely used summarizing measure, the simple average of the survey participants. In particular, the selection process implemented by the SPLS method provides a good prediction performance. The final SPLS forecast combinations are sparse, which implies that forecasters with redundant information or small predictive power over the target are removed from the panel, reducing the estimation error and, therefore, improving the forecast efficiency.

Considering the four subsamples analyzed for the whole period 1991-2012, including the extended sample 2, we find that the SPLS model outperforms the alternative methods in almost 75% of the cases. It is important to highlight that the extended sample 2 includes specifically the crisis period 2008-2009. According to the robustness of the results obtained in the different subsamples and the detailed results for the fourth subsample, the model was able to capture the changes in the behavior of the variables resulting from the financial crisis and performed well during its development, with the exception of HOUSING and BOND. We consider that the prolonged stagnation of housing sector, the severity of the crisis and the sluggish economic recovery have influenced the deterioration of the performance of the forecasting models of these latter variables with respect to the average of forecasters.

Regarding the number of components, the best performing models have just one component and a high degree of sparsity. However, the number of panelists selected depends on the variable and the forecast sample.

The economic uncertainty caused by the financial crisis can be evidenced in the change that occurs in the number of panelists who are selected by the model over the prediction period 2005-2009. For NGDP and UNEMP the number of the panelists selected to build the final combination increased during the recession period 2008-2009. For HOUSING a similar behavior was observed temporarily in accordance with the sector cycle and was repeated, with less intensity, during the recession of the economy.

Further Research

The growing interest in the data rich environment analysis pointed out several extensions to future research such as the construction of models subject to structural breaks, forecasting with cointegrated time series, models with nonlinearities and alternative regularizations methods, among others.

The conflicting results about the stability of the factor loadings and the influence of breaks on the number and estimation of factors, mentioned in the literature review, constitute an important line for future research. Several authors agree on the existence of a structural break after 1985, associated with the Great Moderation (D'Agostino and Giannone, 2006; Breitung and Eickmeier, 2011, among others). For some of them the performance of factor models seems to have lost their advantage in the post 1985 period. However, the authors do not agree on the effects that structural breaks produced in the factor loadings and in the factor estimation. For example while Bates et al. (2012) found that instability in the factor loadings has a limited impact on estimation of the factor space, Breitung and Eickmeier (2011) pointed out that this can cause inconsistent estimates of the loadings and lead to a larger dimension of factor space.

Recent research has been focused on the formulation of test to prove the presence of breaks, but should be extended to the estimation of the impacts that a break could generate in the factor estimation and to methods that take into account in the estimation process the possible existence of structural breaks. Some work in this direction has already been made; see, for instance, Barnejee et al. (2009).

Other area of increasing interest in the literature is regularization; several alternatives have been suggested but it remains an open issue. For example, Allen et al. (2012) proposed some alternative PLS regularization methods such as a Non-negative PLS and Generalized PLS in the context on chemometrics. Croux and Exterkate (2012) proposed a "Robust and Sparse Factor Modelling", a method that combines the robust estimation methods from Maronna and Yohai (2008) and the penalization technique introduced by Witten et al. (2009).

Dissemination of Results

The list of congresses, seminars and publications related to some of the chapters of this thesis is provided below:

Journal Publications

Fuentes, J., Poncela, P. and J. Rodríguez (2014). "Sparse Partial Least Squares in Time Series for Macroeconomic Forecasting". *Journal of Applied Econometrics,* forthcoming. Early View: http://onlinelibrary.wiley.com/doi/10.1002/jae.2384/abstract.

Working Papers

Fuentes, J., Poncela, P. and J. Rodríguez (2012). "Sparse Partial Least Squares in Time Series for Macroeconomic Forecasting". Working Paper 12-22, Statistics and Econometrics Series 16, Universidad Carlos III de Madrid.

Fuentes, J., Poncela, P. and J. Rodríguez (2012). "Sparse Partial Least Squares in Time Series for Macroeconomic Forecasting". Documento de Trabajo 2012-02, Banco Central de Reserva de El Salvador.

Fuentes, J., Poncela, P. and J. Rodríguez (2014). "Selecting and Combining Experts from Survey Forecasts". Working Paper 14-09, Statistics and Econometrics Series 05, Universidad Carlos III de Madrid. This paper has been submitted to an international journal (JCR) for evaluation.

Conferences

- The 30th International Symposium of Forecasting (IFS, 2010). San Diego, United States, 20-13, June. "Sparse Methods for Factor Forecasting: A Comparison".
- The 31st International Symposium of Forecasting (IFS, 2011). Forecasting in a Disruptive World. Prague, Czech Republic, 26-29 June. "Sparse PLS for Macroeconomic Forecasting".
- The 7th International Conference on Computational and Financial Econometrics (CFE, 2013). Senate House, University of London, United Kingdom, 14-16 December. "Multivariate Sparse Partial Least Squares for Macroeconomic Forecasting".

Seminars

- Seminario de Investigación. Departamento de Análisis Económico: Economía Cuantitativa. Universidad Autónoma de Madrid. Madrid, Spain, January 2011. "A Comparison of Sparse Methods for Factor Forecasting".

- Red de Investigadores del Banco Central de Reserva (REDIBACEN). San Salvador, El Salvador, January 2013. “Modelos Factoriales para el Pronóstico de Variables Macroeconómicas”.

References

Allen G.,Peterson, C., Vannucci, M. and Savatic M. (2012). “Regularized Partial Least Squares with an Application to NMR Spectroscopy”. Cornell University.

Banerjee, A., Marcellino, M. and I. Masten (2009), “Forecasting Macroeconomic Variables Using Diffusion Indexes in Short Samples With Structural Change,” in H. Beladi and E. Kwan Choi (eds.), *Forecasting in the Presence of Structural Breaks and Model Uncertainty (Frontiers of Economics and Globalization,Volume 3)*, Emerald Group Publishing Limited: 149-194.

Bates, B., Plagbor-Moller, M., Stock, J. and M. Watson (2012). “Consistent Factor Estimation in Dynamic Factor Models with Structural Instability”.

Breitung, J. and S. Eickmeier (2011). “Testing for Structural Breaks in Dynamic Factor Models”, *Journal of Econometrics*, 163(1): 71-84.

Croux, C., and P. Exterkate (2012). “Robust and Sparse Factor Modelling”. Faculty of Business and Economics. Katholieke Universiteit Leuven.

D’Agostino and Giannone (2006). “Comparing Alternative Predictors Based on a Large Panel Factor Models” Centre for Economic Policy Research, CEPR Discussion Papers No. 6564.

Maronna, R. and V. Yohai (2008). “Robust Low-Rank Approximation of Data Matrices with Element Wise Contamination”. *Technometrics*, 50:295–304.

Witten, D., Tibshirani, R. and T. Hastie (2009). “A Penalized Matrix Decomposition, with Applications to Sparse Principal Component Analysis and Canonical Correlation Analysis”. *Biostatistics*, 10:515–534.